UNIVERZA V LJUBLJANI
SKUPNI INTERDISCIPLINARNI PROGRAM DRUGE STOPNJE
KOGNITIVNA ZNANOST

V SODELOVANJU Z UNIVERSITÄT WIEN,
UNIVERZITA KOMENSKÉHO V BRATISLAVE
IN EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Tine Kolenik

Computer modelling of the influence of natural selection on perceptual veridicality

Računalniško modeliranje vpliva naravnega izbora na veridičnost zaznavanja

MAGISTRSKO DELO

Ljubljana, 2018

UNIVERSITY OF LJUBLJANA
MIDDLE EUROPEAN INTERDISCIPLINARY MASTER’S
PROGRAMME IN COGNITIVE SCIENCE

IN ASSOCIATION WITH UNIVERSITÄT WIEN,
UNIVERZITA KOMENSKÉHO V BRATISLAVE AND
EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Tine Kolenik

Computer modelling of the influence of natural selection on perceptual veridicality

MASTER’S THESIS

Supervisor: Univ. Prof. Dr. Urban Kordeš
Co-supervisor: Univ. Prof. Dr. Igor Farkaš

Ljubljana, 2018

UNIVERZA V LJUBLJANI
SKUPNI INTERDISCIPLINARNI PROGRAM DRUGE STOPNJE
KOGNITIVNA ZNANOST

V SODELOVANJU Z UNIVERSITÄT WIEN,
UNIVERZITA KOMENSKÉHO V BRATISLAVE
IN EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Tine Kolenik

Računalniško modeliranje vpliva naravnega izbora na veridičnost zaznavanja

MAGISTRSKO DELO

Mentor: izr. prof. dr. Urban Kordeš
Somentor: univ. prof. dr. Igor Farkaš

Ljubljana, 2018

ACKNOWLEDGEMENTS

To *my parents* for the eternal support and for genuinely believing that knowledge is a gift no one can take away from you. This thesis is dedicated to you.

To *Polona* for making sure I did not go (too) crazy while writing this thesis.

To Professor *Urban Kordeš* for the continued inspiration and true mentorship.

To Professor *Igor Farkaš* for the help and the always uplifting words.

To *the "Observatory" research team* for the ideas.

ZAHVALA

Staršem za večno podporo in pristno vero v rek, da ti znanja ne more vzeti nihče. Magistrsko delo je posvečeno vama.

Poloni, ker si poskrbela, da se mi med pisanjem magistrskega dela ni (še bolj) zmešalo.

Profesorju *Urbanu Kordešu* za nenehen navdih in resnično mentorstvo.

Profesorju *Igorju Farkašu* za vso pomoč in vedno vzpodbujajoče besede.

Skupini »Observatorija« za vse ideje.

Abstract

The thesis explores computer modelling and its value in cognitive science as natural epistemology. This exploration is realised on several levels of analysis in terms of abstractness. Cognitive science and epistemology are argued to be closely related, manifesting their overlap in what some proponents of this connection name natural epistemology. The latter is defined as the study of epistemological questions with scientific methods. Key elements of natural epistemology are identified and proposed, most importantly the loop of knowings between epistemological insights in cognitive science and epistemology of cognitive science, which characterises progress in natural epistemology. Cognitive science and epistemology both primarily wonder what the relationship between mind and world is, and perception is identified as one of the sources on the knowledge of the world. It is therefore chosen for investigating this relationship, taking the evolutionary perspective on the development of perception. Computer modelling with genetic algorithms is used to study whether it is isomorphic or non-isomorphic perception that is more beneficial for modelled organisms. Two computer models are introduced – a model presented by Donald D. Hoffman and his colleagues, which possesses cognitivist presuppositions, and a newly designed model, which builds on Hoffman's model by replacing certain cognitivist presuppositions for enactivist ones, mostly focusing on the addition of a sensorimotor loop. The models both produce the same results, as they show that non-isomorphic perception is evolutionary more beneficial than isomorphic. However, the sensorimotor loop causes the newly designed model to evolve faster. Afterwards, computer modelling is presented in the light of cognitive science as natural epistemology, questioning the results' validity. The value and role of computer modelling is shown to be historically monumental by placing it in the loop of knowings and showing its influence on epistemological insights in cognitive science as well as epistemology of cognitive science. Despite the influence, several problems are identified, especially the "PacMan Syndrome", the problem of the designed agents being unable to self-determine their meaning, which is forced upon them by the designer instead. The value of the two implemented models is discussed in this light. Two essential questions are posed: What do the models tell us about cognition? What role does their modelling play, especially the approach of designing models with different (epistemological) presuppositions and discerning their influence on final results? The first question is addressed by evaluating the models in several areas. The models are found to be explanatory of a possibility of non-isomorphic perception evolving (as opposed to the prevalent thoughts on that not being possible), not predictive (as they are not meant to be), abstract and simple, which may hinder the approach of comparing the models for their presuppositions, as they might not be able to affect the results because of the simplicity. Regarding the role of genetic algorithms, their arbitrariness in certain elements is presented as problematic, but as even more problematic, the design of the fitness function is presented. The fitness function is identified as an instantiation of the PacMan Syndrome, as the fitness function dictates what is good and what is bad for the models' agents. It is suggested that by making the fitness function evolvable phylogenetically and ontogenetically, the designer's role in predictably forcing its own meaning onto the agent is diminished a bit. By making the models more complex, the approach of comparing them would be made more legitimate in this case, but it was argued that it was a useful approach, as it showed the value of the sensorimotor loop. Regarding the models' value on learning about cognition, it is suggested that they offer a functional understanding of a possible occurrence of non-isomorphic perception. Finally, the models are placed in the loop of knowings, their possible influence speculated upon.

Keywords: cognitive science, computer modelling, enactivism, epistemology, evolution, genetic algorithms, perception

Povzetek

Magistrsko delo raziskuje računalniško modeliranje in njegovo vlogo v kognitivni znanosti kot naravni epistemologiji. Raziskovanje se odvija na različnih nivojih abstraktnosti. Kognitivna znanost in epistemologija sta predstavljeni kot sorodni disciplini pod imenom naravna epistemologija, ki označuje raziskovanje epistemoloških vprašanj z znanstvenimi metodami. Predstavljeni so ključni elementi naravne epistemologije s poudarkom na zanki med védenjem znotraj kognitivne znanosti in védenjem kognitivne znanosti. Razmerje med zunanjim svetom in umom predstavlja eno glavnih tem kognitivne znanosti in epistemologije, zaznavanje pa je opredeljeno kot eno glavnih virov za spoznavanje zunanjega sveta in zato zanimivo za raziskovanje. Evolucijski vidik je pokazan kot en najzanimivejših obravnavanj zaznavanja. Računalniško modeliranje z genetskimi algoritmi je uporabljeno za raziskovanje vprašanja, ali je za modeliran organizem preživetveno koristnejše izomorfno ali neizomorfno zaznavanje. Predstavljena sta dva modela – kognitivistični model Donalda D. Hoffmana in sodelavcev ter lasten model, ki nekatere kognitivistične predpostavke modela Hoffmana in sodelavcev zamenja z enaktivističnimi, s poudarkom na senzomotorični zanki. Oba modela rezultirata v razvoju neizomorfnega zaznavanja kot preživetveno koristnejšega za modelirane organizme. Senzomotorična zanka v enaktivističnem modelu se izkaže za koristno, saj povzroči hitrejši razvoj v modelu. Po predstavitvi modelov magistrsko delo razišče vlogo računalniškega modeliranja za naravno epostemologijo, predvsem z namenom prevpraševanja rezultatov predstavljenih modelov. Vloga računalniškega modeliranja je prestavljena kot zgodovinsko izredno pomembna v kognitivni znanosti, kar se kaže, ko je metoda postavljena v zanko védenj. Kljub pomembnosti je izpostavljenih več težav, predvsem Pacmanov sindrom, ki označuje težavo od raziskovalke vsiljenega pomena v modeliranih agentih, ki se ne morejo samodoločati. V tej luči se glede uporabnosti implementiranih modelov postavljata dve vprašanji: Kaj nam modela povesta o kogniciji? Kakšno vlogo igra modeliranje, še posebej pristop primerjanja vloge različnih (epistemoloških) predpostavk v samih modelih? Prvo vprašanje je naslovljeno z vrednotenjem modelov na več področjih. Ugotovljeno je, da modela pojasnjujeta možnost razvoja neizomorfnega zaznavanja (ki gre proti prevladujoči ideji, da je evolucija takšnega zaznavanja nemogoča) in ne napovedujeta (saj temu nista namenjena). Ugotovljeno je tudi, da sta abstraktna in preprosta, kar oteži pristop primerjave modelov na podlagi njihovih predpostavk, saj se zdi, da te zaradi preprostosti ne morejo vplivati na končni rezultat. Raziskana je vloga genetskih algoritmov, v katerih sta odkriti težavi v njihovi delni arbitrarnosti ter v predpostavljeni kriterijski funkciji. Ta je označena kot primer Pacmanovega sindroma, saj kriterijska funkcija veluje, kaj je za organizem koristno in kaj nekoristno. Predlagano je, da se raziskovalkina moč v določanju razvoja agentov lahko omeji z oblikovanjem kriterijske funkcije tako, da se le-ta filogenetsko in ontogenetsko razvija. S tem se moč ustvarjalke modeliranih agentov v tem, da lahko namerno določa in vnaprej pozna njihov razvoj, omeji. Pristop primerjave modelov na podlagi njihovih predpostavk in vpliva le-teh na rezultate je spoznan za legitimnega, z opozorilom, da sta implementirana modela morda premalo kompleksna, saj se zdi, da predpostavke ne morejo vplivati na končni rezultat. O kogniciji implementirana modela povesta to, kako se neizomorfno zaznavanje lahko potencialno razvije, na koncu pa sta modela uvrščena še v zanko o védenju znotraj kognitivne znanosti in védenju kognitivne znanosti, kar nudi razmislek o njunem možnem vplivu.

Ključne besede: enaktivizem, epistemologija, evolucija, genetski algoritmi, kognitivna znanost, računalniško modeliranje, zaznavanje

Contents

0 Introduction

The question of the relationship between mind and world is a question without a definitive answer. The burden of answering it has been, to different lengths, carried by a multitude of disciplines, resulting in many views and theories, yet no meaningful consilience has been reached. Philosophical epistemology (Kvanvig, 2003) has developed as the foremost to carry the torch of knowledge on this enigma, but the advent of cognitive revolution in science has introduced the scientific method to the extensive groundwork on the many varieties and off-shots of what it is to know. The scientific structure researching knowing that has spawned from a wide interdisciplinary field of disciplines such as artificial intelligence, psychology, neuroscience, linguistics and so on, with foundations in philosophy, organised its workings and advancements in a peculiar way, where the bottom-up knowledge from researching how living beings cognise influenced the nature and the progress on the question of the relationship between mind and world as much as the top-down researchers' presuppositions on how living beings cognise. The intertwinings of the two has defined many of the breakthroughs on knowing in science, yet the epistemological question is still looming on the horizon. Many researchers have proposed that computer modelling and artificial intelligence may help solving such challenging theoretical disputes (Froese, 2007). Perception, being one of the foremost sources of knowledge of the external world (Alston, 1999), has been of a distinct interest of a large body of research in various disciplines, where one of the most interesting perspectives of investigating perception in relation to the knowledge of the external world has been the evolutionary perspective. Computer modelling has been used to gauge the evolution of perception and how the perception mediates between the external world and the mind (Hoffman, Prakash, & Singh, 2015).

The thesis explores the evolution of perception and whether natural selection shapes cognition into representing the external world as it is or not, asking what kind of perception has more survival value for organisms. This is explored with computer modelling, specifically with genetic algorithms (Mitchell, 1999). However, the implemented models serve mainly as a means to address grander issues, namely the background of such modelling, how useful computer modelling is as a method for investigating the relationship between mind and world, and what can be discerned from the implemented models about cognition (Riegler, Stewart, & Ziemke, 2013). The topics are tackled by investigating computer modelling in general as well as examining the models and their results implemented for this thesis. Computer modelling of the influence of natural selection on perceptual veridicality is therefore shown to be a much more complex endeavour than it might appear at first sight.

Since the thesis covers a wide net of approaches, ideas, levels of explanations and yet still tries to weave a cohesive narrative, the scheme below can be used for orientation. The scheme will be present at specific points of the thesis, showing the location in the narrative. The narrative follows the shape of an hourglass, with the general topics being discussed at the top or the beginning (cognitive science, epistemology), then zooming in and addressing more and more specific issues when going down the hourglass towards the neck (evolutionary perspective on perception) with the nexus in the neck being the empirical part of the thesis (computer models), then widening the scope again when entering the second, bottom bulb by discussing the more general topics by applying the knowledge gained from hands-on computer modelling (the role of models in general and the implemented models for the thesis in investigating epistemological questions).

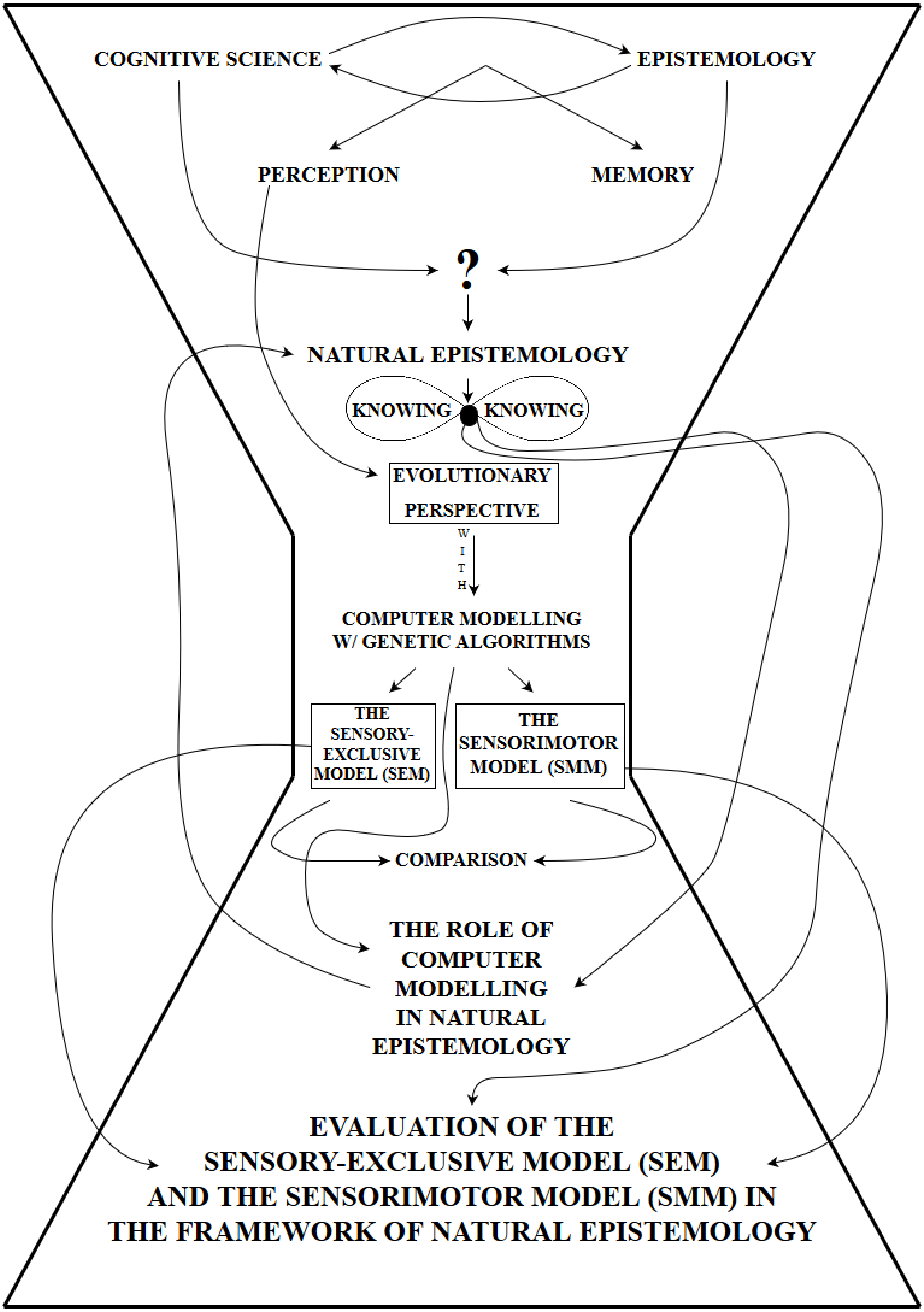

Figure 0. The schematic for navigating the thesis.

1 Cognitive Science as Natural Epistemology

The first chapter of the thesis will be dedicated to presenting a premise that characterises cognitive science as natural epistemology. The process of presenting the premise will consist of slowly building the case for cognitive science being imagined as natural epistemology – by showcasing similarities between cognitive science and epistemology, their historical relationship, their overlap, discussing examples where they seem to collide, etc. – as well as trying to discern some of the internal mechanics that make cognitive science work as natural epistemology. The latter will be shown through presenting the progress of cognitive science in terms of epistemological shifts that work as a loop between two important concepts – knowing within cognitive science and knowing of cognitive science.

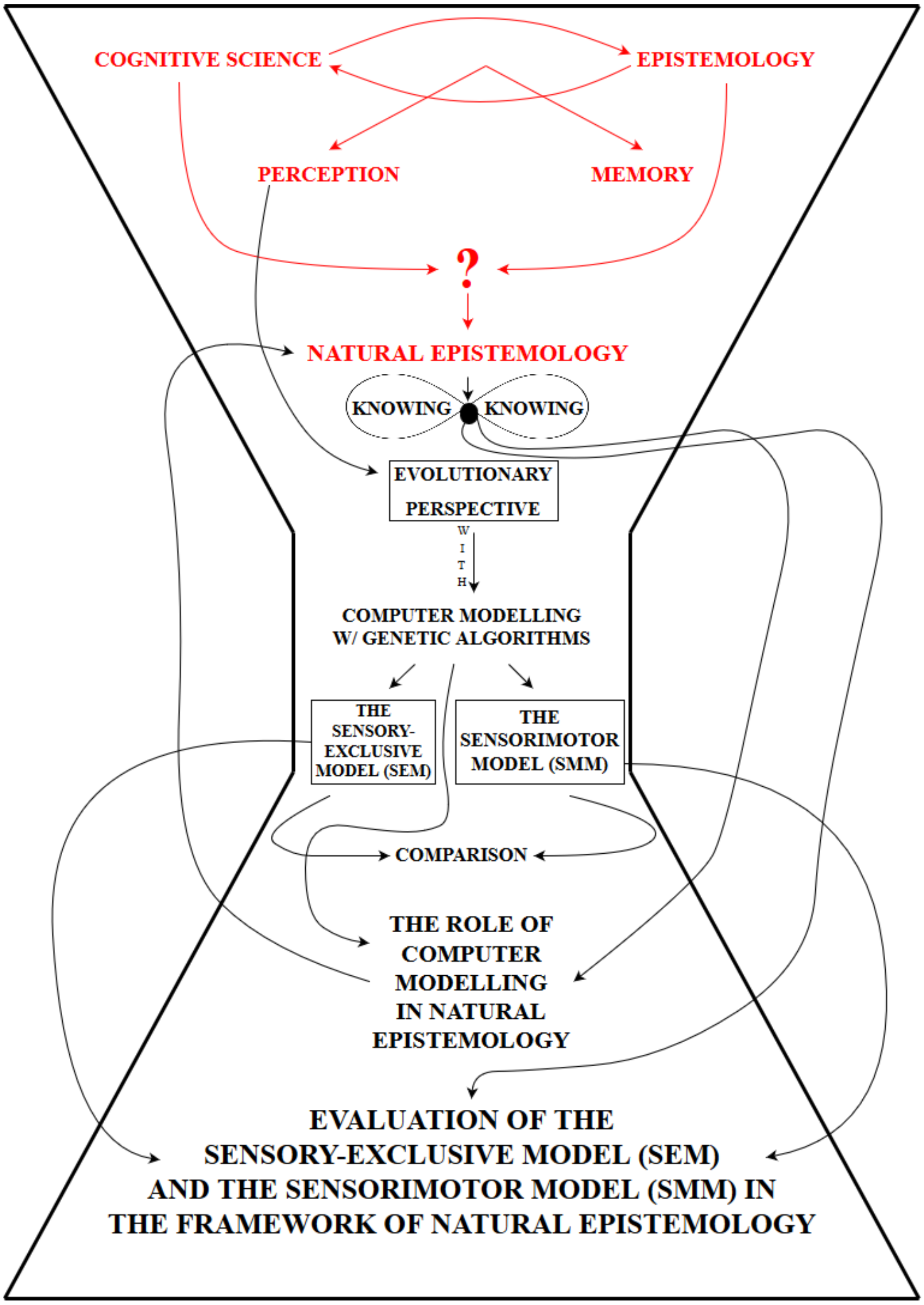

Figure 1. The position in the narrative schematic of the thesis, marked in red.

1.1 Epistemology and Cognitive Science: One, but Not the Same?

Is the brain in a vat, a common thought experiment device for showing the strenuous relationship between knowledge, reality and the mind, only of philosophical value or does it have scientifically-relevant upshots as well? Imagine having an isolated brain in a laboratory setting, where the brain is treated with processes-sustaining procedures to keep it active, while at the same time being fed with electrical impulses in the same manner as the embodied brain is. This is a common scenario for aforementioned thought experiments, where the person of such a disembodied brain would consciously perceive as all other, embodied people, and would hold the belief that she is experiencing the real world. The usual thought experiment where this scenario is used is about not being able to disassociate the real world from a simulation of it, thus pointing to a philosophical school of thought called scepticism, where the certainty of knowing something is always under question (Klein, 2015). But when the brain in a vat is reinterpreted and probed in different manners, one can arrive at a place where epistemology and cognitive science seem to collide. Since the disembodied person is supposedly experiencing what an embodied person experiences, this begs the question of how the senses give rise to experience and therefore knowledge of the world (Alston, 1999). Is wondering about the sources of knowledge, and how these sources, namely perception, give rise to knowledge of the world, a philosophical (in the domain of epistemology) or scientific (in the domain of cognitive science) endeavour? Is wondering whether different embodiments (and let disembodiment be one such form of embodiment) give rise to different "knowings" of the world a philosophical or a scientific endeavour?

The brain in a vat represents an exemplary gateway to acknowledging that certain questions, present in epistemology for centuries, have a particular character where they can be construed as scientifically engaging, especially for cognitive science. This is especially true as it is not inconceivable that in a few decades such a thought experiment, which is a frequent method in philosophy to think through a hypothesis or a theory and reach its consequences, could actually be empirically feasible. Scientific advancement in cognitive science can make natural scientific methods the next stage for theories that were before treated with philosophical methods; therefore, philosophical epistemology could lead straight into empirical cognitive science (and after that, back to philosophy again, forming a loop). There are polar opposite stances on cognitive science being able to elucidate philosophical concepts. On the one end of the spectrum, Peter Hacker, an Oxfordian philosopher, maintains that philosophical inquiry is an entirely different endeavour than scientific inquiry, saying that philosophy "is not a contribution to human knowledge, but to human understanding" (Hacker, 2001, p. 141). He rejects experimental constituents of cognitive science as being able to investigate the mind as an extension or the next step of philosophy, going so far as to calling neuroscience nonsensical (Garvey, 2010). On the other end of the spectrum, Patricia S. Churchland believes that cognitive science, specifically neuroscience, will eventually replace philosophy (Churchland, 1986), which follows the footsteps of the foremost Enlightenment philosopher John Locke, who regarded philosophy as a handmaid of science (Ciulla, 2011).

Regardless of the extreme positions on the role of cognitive science in relation to philosophy, the hold of cognitive science on epistemology goes beyond mere possibility. Even the first's name gives it away – to cognise, after all, means to know, to understand (Merriam-Webster.com, 2018). Basic descriptions of cognitive science, featured on college or encyclopaedia websites, directly mention knowledge or knowing as the central theme of the research field. Various phrasings, such as "the broad goal of cognitive science is to characterize the nature of human knowledge" ("Cognitive Science", n.d., para. 1), "over the years it has become apparent that people in fields such as philosophy, psychology, linguistics, computing science, and neuroscience — among others — have been asking essentially the same questions […]: What do we know, and how do we know it?" ("Cognitive Science At Simon Fraser University," n.d., para. 2) and "a central epistemological question [of cognitive science] is how minds gain knowledge of the external world" (Thagard, 2013, para. 3) coalesce into a very similar if not the same meaning. However, only acknowledging this possible overlap between epistemology and cognitive science may not be enough to make a case for a deeper connection. It is also not enough to take a common formulation of what epistemology is and how literally the same can be said for cognitive science (or at least the part of cognitive science that deals with direct cognising-related questions). Even so, Kvanvig (2003, p. ix) has this to say about epistemological endeavours: "Philosophers reflect on the nature and extent of knowledge not simply because they have free afternoons to fill but (also) because questions about what we know and how we know it touch on the deeply significant questions about the relationship between mind and world and the possibility of success in determining what is true and what is not." For contrast, this is how Ó Nualláin (2002, p. 4) describes cognitive science: "Cognitive Science is a discipline with both theoretical and experimental components which, inter alia, deals with knowing. In doing so, it quite often finds itself walking in the footprints of long-dead philosophers, who were concerned with the theory of knowledge (epistemology)." When writing about the struggles of cognitive science, he continues (Ibid., p. 5): "[…] the struggle […] was that with the more general problem of knowledge. The lines of approach taken to this problem were extremely varied. The key to the myriad conceptions of knowledge which arose is consideration of the problem of the relationship between mind and world." Kvanvig and Ó Nualláin both use the same phrase when describing their subject matter – the relationship between mind and world. They both characterise knowledge or knowing as their subject matter's foundation. Even more so, epistemology is directly mentioned by Ó Nualláin as the main source of investigative matter for cognitive science. However, a few questions loom over this entire issue, which should be considered as not to trivialise the connection between epistemology and cognitive science: Are cognitive science and epistemology really asking the same questions, and if so, which ones? Are epistemological questions really accessible to natural methodology of (cognitive) science? Are there accounts of historical relations between epistemology and cognitive science? Is there existing propagation of the idea that epistemology and cognitive science are or should be significantly related?

The introduction of epistemology and cognitive science as something that could be seen in a similar light serves the explication of my proposal that will be fleshed out in this chapter. I propose that there exists a consequential overlap between cognitive science and epistemology, as presented in Figure 2. At this point, I am hesitant to making strong proclamations of what this overlap is; this is treading precarious grounds where it would be disingenuous making too many definitive statements. However, this overlap – what may be in it, how extensive it is, etc. – will be explored in the present chapter.

The deeper connection between epistemology and cognitive science will be demonstrated by examining individual epistemological questions, especially as articulated by prominent philosophers. The examined epistemological questions will then be connected to research in cognitive science and its constituent fields by looking at answers that these fields may offer. This endeavour appears to be a lot like writing a historical account of philosophical roots of cognitive science in general. Two phenomena will be more thoroughly explored: perception and memory. The justification for choosing the two phenomena and several accounts of how epistemology and cognitive science ask and answer questions relating to these two phenomena as epistemological concepts will follow. After that, a few examples will be listed illustrating that such relationships have been noted by other authors as well, solidifying the notion of the deep relationship between epistemology and cognitive science by showing that it has already been explicated. This historical delve will serve as an opening to discuss how much epistemology and cognitive science overlap. In the end, all the build-up of exploring the relationship between epistemology and cognitive science will be summarised, revealing an emergent area of research.

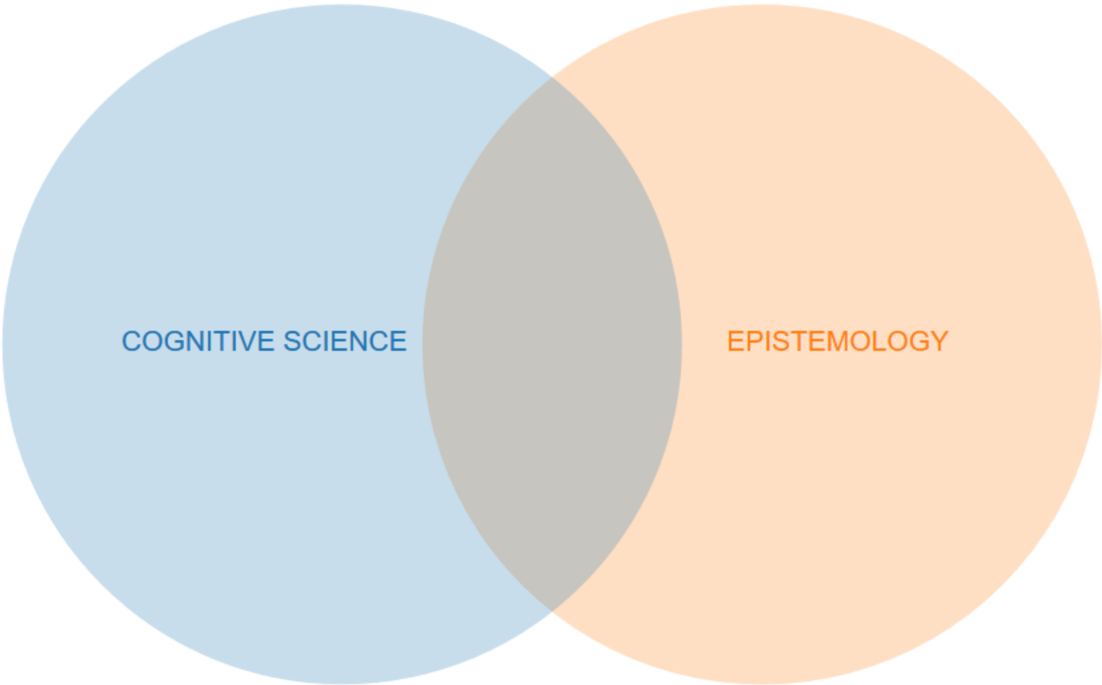

Figure 2. A seemingly possible overlap between questions in cognitive science and epistemological questions in philosophy.

1.1.1 Perception

In epistemology, perception is regarded as one of the sources of knowledge of the outside world (Alston, 1999). Perception, especially visual perception, will be a recurring topic in this thesis, so it is appropriate for it to be thoroughly examined.

In his criticism against material objects, Berkeley (1710/1982) makes this argument:

> They who assert that figure, motion, and the rest of the primary or original qualities do exist without the mind in unthinking substances, do at the same time acknowledge that colours, sounds, heat cold, and suchlike secondary qualities, do not--which they tell us are sensations existing IN THE MIND ALONE [all-caps Berkeley], that depend on and are occasioned by the different size, texture, and motion of the minute particles of matter. This they take for an undoubted truth, which they can demonstrate beyond all exception. Now, if it be certain that those original qualities ARE INSEPARABLY UNITED WITH THE OTHER SENSIBLE QUALITIES [all-caps Berkeley], and not, even in thought, capable of being abstracted from them, it plainly follows that they exist

> only in the mind. But I desire any one to reflect and try whether he can, by any abstraction of thought, conceive the extension and motion of a body without all other sensible qualities. For my own part, I see evidently that it is not in my power to frame an idea of a body extended and moving, but I must withal give it some colour or other sensible quality which is ACKNOWLEDGED [all-caps Berkeley] to exist only in the mind. In short, extension, figure, and motion, abstracted from all other qualities, are inconceivable. Where therefore the other sensible qualities are, there must these be also, to wit, in the mind and nowhere else. (Berkeley, 1710/1982, p. 13)

Similarly, questioning our perception of the world and how we construe the latter can be discerned in artificial intelligence, when cognitive scientists hit a wall with their computer vision research. Imagine circling a kitchen table, your perspective on it always changing. How do we know we are always looking at the same object, since the image is always different? The approach taken in computer vision has been that of continually updating the image and comparing it with stored images of the table (Marr, 1982). However, this is extremely time consuming, memory hogging and therefore ecologically unviable. It is also one of the scenarios that has led to the articulation of the frame problem, "the challenge of representing the effects of action in logic without having to represent explicitly a large number of intuitively obvious non-effects" (Shanahan, 2016, para. 1), which has caused major setbacks in artificial intelligence research.

Both, epistemological and cognitive-scientific inquiries on perception, are centred on the problem of perceptual knowledge, specifically on the questions of what the role of causation in perception is, how perception gives rise to experience, where the concept of what we are perceiving comes from, and how this concept is connected to perception and experience of it as well as the (sometimes presupposed) outside objects of it. Computer vision, sometimes inadvertently, tests philosophical ideas on perception, and presents answers through its own methods. The connection between epistemological thoughts on perception (in this case, specifically by Berkeley), and more contemporary work on perception in computer vision and cognitive science, present a case for complementary enterprise where the two are seen not as separate, but as one, with a striking difference – the methodology. Berkeley's argument against material objects and his treatise of perception wagers on philosophical analysis, while computer vision in artificial intelligence follows the creed of "understanding by [designing and] building" (Pfeifer & Scheier, 1999, p. 22), a very hands-on approach of synthetic approach to science, which Mirolli and Parisi (2011) distinguish from the analytic approach to science – the synthetic approach comprises of computer and robot models to research phenomena, while the analytic approach comprises of observation and experimentation to research phenomena.

1.1.2 Memory

Steup (2005) lists, alongside perception, four more sources of knowledge, considered in epistemology: introspection, memory, reason and testimony. About memory, Steup writes:

> One issue about memory concerns the question of what distinguishes memorial seemings from perceptual seemings or mere imagination. Some philosophers have thought that having an image in one's mind is essential to memory, but that would appear to be mistaken. When one remembers one's telephone number, one is unlikely to have an image of one's number in one's mind. The […] questions about memory are these: […] what makes memorial seemings a source of justification? […] how can we

> respond to skepticism about knowledge of the past? Memorial seemings of the past do not guarantee that the past is what we take it to be. (Steup, 2005, para. 88)

Cognitive science's exploring of memory is rich and comprehensive, and distinguishing memories from other similar phenomena is at the forefront (Schacter et al., 2012). There is fMRI research in differences in memory and imagination in the brain (Kirwan, Ashby, & Nash, 2014), where the studies "have shown that remembering and imagining utilize the same neural substrates including the hippocampus, and are therefore intricately related" (Ibid., p. 1), but it was recently discovered that "while the hippocampus seems to be involved in both remembering the past and imagining the future, the pattern of activity within the hippocampus distinguishes between these two different tasks" (Ibid., p. 1). There is research on our ability to evaluate whether memories reflect real or imagined events. For example, interpersonal reality monitoring refers to evaluating others' memories. Research (Clark-Foos, Brewer, & Marsh, 2015, p. 427) shows that "people are better than chance and that the ability to accurately make this judgement can be improved or reduced with appropriate and inappropriate training, respectively." Steup's question about scepticism about knowledge of the past refers to another leading problem in the science of memory, namely the one about memory's accuracy. Daniel Schacter, the leading researcher on memory and former chair of Harvard University's Psychology Department, wrote a comprehensive book on memory's fallibility, aptly named *The Seven Sins of Memory: How the Mind Forgets and Remembers* (2001). In it, he lists seven memory's features he characterises as sins:

- transience, denoting the general deterioration of a specific memory over time (e.g., memories in further past are less accessible),
- absent-mindedness, denoting blank spots in memory because of insufficiently paid attention at the time of the event (e.g., forgetting whether the door has been locked),
- blocking, denoting the phenomenon where another memory or piece of information interferes with another one or "stepping in" in its place (e.g., to have a word at the tip of the tongue),
- misattribution, denoting the false memory of a source of specific information (e.g., pointing to the wrong suspect from a police line-up just because they are there),
- suggestibility, denoting the context in which a memory is remembered, where it can be influenced by suggestive participants (e.g., leading questions in the courtroom),
- bias, denoting the plasticity and reconstructive nature of memories when it comes to being influenced by personal worldview, knowledge, emotions, etc. (e.g., racial biases),
- persistence, denoting consciously imposing memories, which can lead to a wholly transformative experience in personality and thus stimulating most, if not all the listed sins so far (e.g., post-traumatic stress disorder).

Epistemological questions concerning memory and its relation to reality and imagination play a significant role in cognitive science and its research about memory. The question of "what distinguishes memorial seemings from perceptual seemings or mere imagination" (Steup, 2005, para. 88) has been tackled with, e.g., imaging methods (Kirwan et al.'s research), while questions of scepticism about knowledge of the past fuelled whole books of answers, produced with scientific methods (Schacter's book). As before with perception, it certainly appears as if cognitive science does offer at least some kind of answers to epistemological questions, especially on perception and memory, and these epistemological questions bear a striking resemblance to questions posed by cognitive scientists who work on the two phenomena.

In this section, I tried to make a deeper connection between epistemology and cognitive science by looking at specific epistemological questions, namely on perception and memory, and by presenting a case that they are similarly posed in cognitive science. The latter does produce answers through its methods, which can be construed at least as partial, scope-specific answers to epistemological questions as such. This deeper connection has also been noted by other authors, which I will explore in the next section.

1.1.3 Historical Relationship between Epistemology and Cognitive Science

The close relationship between epistemology and cognitive science (or its constituent disciplines) has been noticed by a number of authors, not only by, e.g., Ó Nualláin. Many studies can be found on these telling pairings between epistemological and scientific inquiries. Fabricius (1983) documents a remarkable similarity between Immanuel Kant’s and Jean Piaget’s work on the question of what knowledge is and how it develops, where it is again the methods that diverge[1]. Hawkins (2011) provides a concise history of psychophysics – William James, Ernst Heinrich Weber, Gustav Fechner, Hermann von Helmholtz, Wilhelm Wundt, etc. – and their endeavour for sketching “a new sort of epistemology, explaining the reality of the mental and the organic, bridging the cleft that separates nature and consciousness, reality and perceptual appearance, and combining science with direct human experience” (Heidelberger, 2004 in Hawkins, 2011, “Conclusion: Radical Phenomenalism”, para. 3), basing their work on, e.g., Spinoza and Leibniz. Garrett (1999) offers an insight into Skinner’s importance for epistemology:

> In order to evaluate even this indirect and suggestive contribution of Skinner's, it will be helpful to begin by stating the central goal or purpose of epistemology as it is understood by most epistemologists: As responsible thinkers we all want to hold a belief if and only if it is true. The central goal of epistemology is, therefore, to help us distinguish truth from falsity. […] If Skinner's work has any significance for epistemology, therefore, it is most likely to be found in his work on verbal behavior […]. Skinner himself well understood this as the following statement clearly indicates. (Garrett, 1999, p. 69–70)

Garrett (Ibid., p. 70) goes on to quote Skinner (1957): “One of the ultimate accomplishments of a science of verbal behavior may be an empirical logic or a descriptive and analytic scientific epistemology.” He continues: “[…] truth is a concept of central importance to epistemology and […] the most important contribution of Skinner's work to epistemology arises from its implications for the analysis of truth and related concepts […]” (Ibid.). The direct relevance of epistemological questions in psychology, neuroscience and AI, noted not only by myself, but other authors as well, is revealing itself to be one of the most effective glues for various constituent disciplines of cognitive science, making it a legitimate interdisciplinary project.

However, the parallelity in question and dichotomy in answer between philosophical epistemology and scientific epistemology is not limited only to case resemblance and wondering similarities. A substantial number of philosophers and scientists has been calling for and developing an enterprise consisting of the intersection between epistemology and

[1] It is interesting that Piaget's theoretical framework for how knowledge develops has been named genetic epistemology.

science. They have also noticed that there is something significant in this connection that needs examination and, ideally, a model or a distinct discipline. Cognitive science presents itself as a contender, with relatively diverse claims of how much of cognitive science is epistemological as described. Ó Nualláin (2002) is bold in his statements, saying that in "a limited sense, [cognitive science] is and always has been epistemology" (Ibid., p. 5). When describing what philosophical epistemology is, he writes (Ibid., p. 13): "A short answer to this question is that it is the theoretical approach to the study of Knowledge. It can be distinguished, in these terms, from experimental epistemology which features in the remainder of the disciplines within Cognitive Science." Cognitive science, according to Ó Nualláin, can "experimentally test conjectures of [epistemologists], or on occasion […] establish that these conjectures are too abstract to be so tested" (Ibid., p. 4). Ó Nualláin goes very far in his position that cognitive science in general is epistemology. It would certainly not be hard to agree with him that there would be no cognitive science without epistemology as a philosophical field. It can therefore be said that cognitive science investigates knowing of living beings, primarily humans. However, this does not necessarily mean that all undertakings of cognitive science are epistemological.

1.1.4 Epistemological or Not?

Ó Nualláin states firmly that cognitive science is and always has been epistemology. It would not be hard to imagine that there are people, concerning themselves with the overlap between epistemology and cognitive science, that would oppose this. I will remain somewhat agnostic, but I can think of examples that may be construed as not epistemological, as they may, in a sense, not offer "direct" answers to how living beings cognise. To give an example: Adolphs (2015) lists a number of (unsolved) questions in cognitive science and (cognitive) neuroscience, and the majority may be construed as not being epistemological and not directly concerned with the question of knowing. Below is a sample of them (Adolphs, 2015, p. 173–174):

1. "What is the connectome of a small nervous system, like that of Caenorhabitis elegans (300 neurons)? […] What is the complete connectome of the mouse brain (70 000 000 neurons)? […] What is the complete connectome of the human brain (80 000 000 000 neurons)?"
2. "How can we image a live brain of 100 000 neurons at cellular and millisecond resolution? […] How can we image a live human brain at cellular and millisecond resolution?"
3. "How do circuits of neurons compute? […] How does the mouse brain compute? […] How does the human brain compute?"

These questions definitely have the character of not being epistemological, yet belong to the field of cognitive science, as neuroscience is one of its core areas (Bermúdez, 2014). On the other hand, these kinds of questions may represent tiny pebbles in the mosaic of our knowledge of cognition, and may lead to answers to epistemological questions, therefore still being part of the epistemological family that cognitive science seems to be. Ó Nualláin – who is not completely sold on neuroscience being strictly a part of cognitive science (Ó Nualláin, 2002) – uses a peculiar phrase to describe epistemological questions that can undergo empirical testing: experimental epistemology. Following Ó Nualláin's train of thought in this manner, cognitive science could be thought wholly as epistemology, having two distinct branches, philosophical and empirical epistemology. The latter would consist of all constituent disciplines of cognitive science minus philosophy.

As stated, my position on this is a bit agnostic. I acknowledge the substantial overlap between epistemology and cognitive science, but I will not delve into whether questions such as listed neuroscientific ones are epistemological or not – mostly as it is not relevant to this thesis. The more important giveaway at this point is the label used by Ó Nualláin, "empirical epistemology". This is not the first time such a phrase has been used, although it has never been used in such direct association with cognitive science.

1.1.5 Natural Epistemology

Progressing research in epistemology in a pure conceptual way by contemplative philosophical work is definitely a powerful instrument for gaining new knowledge. However, Willard Van Orman Quine, a philosopher and logician, insightfully comprehended what Dennett (1996, p. 134) colourfully described with these words: "Just as you cannot do very much carpentry with your bare hands, there is not much thinking you can do with your bare brain." Quine coined a term very similar to what Ó Nualláin deems as "empirical epistemology". Quine's "naturalized epistemology" describes a view wherein epistemology includes scientific methods:

> Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz. a physical human subject. This human subject is accorded a certain experimentally controlled input - certain patterns of irradiation in assorted frequencies, for instance - and in the fullness of time the subject delivers as output a description of the three - dimensional external world and its history. The relation between the meager input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology; namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence. (Quine, 1969, p. 82–83)

Almost in parallel, equally strong, congruent views on epistemology came from the more scientifically-inclined circles of cybernetics. Gregory Bateson, one of the founders of cybernetics and second-order cybernetics movement, wrote this in his seminal book *Mind and Nature* (1979, p. 32): "The processes of perception are inaccessible; only the products are conscious and, of course, it is the products that are necessary. The two general facts-first, that I am unconscious of the process of making the images which I consciously see and, second, that in these unconscious processes, I use a whole range of presuppositions which become built into the finished image-- are, for me, the beginning of empirical epistemology." Alongside "empirical epistemology" Bateson also uses the term "experimental epistemology" (Ibid., p. 32), which is also what Ó Nualláin uses. Bradford Keeney, another cyberneticist and Bateson's doctoral student, characterises the endeavour, using the term "natural epistemology", as such: "Epistemology emerges from creatura: Even to know that there is a world of no distinctions requires that we draw a distinction. From the perspective of pleroma, all the distinctions we create are illusion or maya, the incomplete side of a more encompassing view in which there are no distinctions. As natural epistemologists, our dilemma is having to draw distinctions in order to know a world, while knowing that these constructions are illusory" (Keeney, 1983, p. 63). Keeney's "epistemology emerges from creatura" may especially hold a clue that predicts cognitive science in a sense described earlier – as a science concerned with knowing of living beings, while Bateson focuses on presuppositions held by these living beings that shape their perception and experience, which may be interpreted as biological substratum, being similar to Piaget's genetic epistemology. Other cyberneticists espousing not only similar views, but also similar terms, include Niklas Luhmann and Humberto Maturana. Luhmann, being influenced by Quine, uses natural epistemology due to "[n]atural consist[ing] in the perception of knowledge as a series of significant operations in an observer and hence in a disregard of the question of true and false" (Thyssen, 2004, p. 8), believing that it is very important to "provide an empirical description of cognitive operations" (Ibid.) to "analyze what has to be presupposed when a system observes" (Ibid.).

A concise definition of natural epistemology, borrowing from all of its proponents above, can therefore be the following: Natural epistemology is the study of epistemological questions with the use of natural scientific methods. This definition helps fill the intersection of the circles in Venn diagram in Figure 2, which is featured in Figure 3. Returning to Kvanvig's and Ó Nualláin's characterisation of epistemology as primarily considering the relationship between mind and world, and therefore knowing, it is important to delve into this notion as dealt with in cognitive science through its evolution as a scientific field.

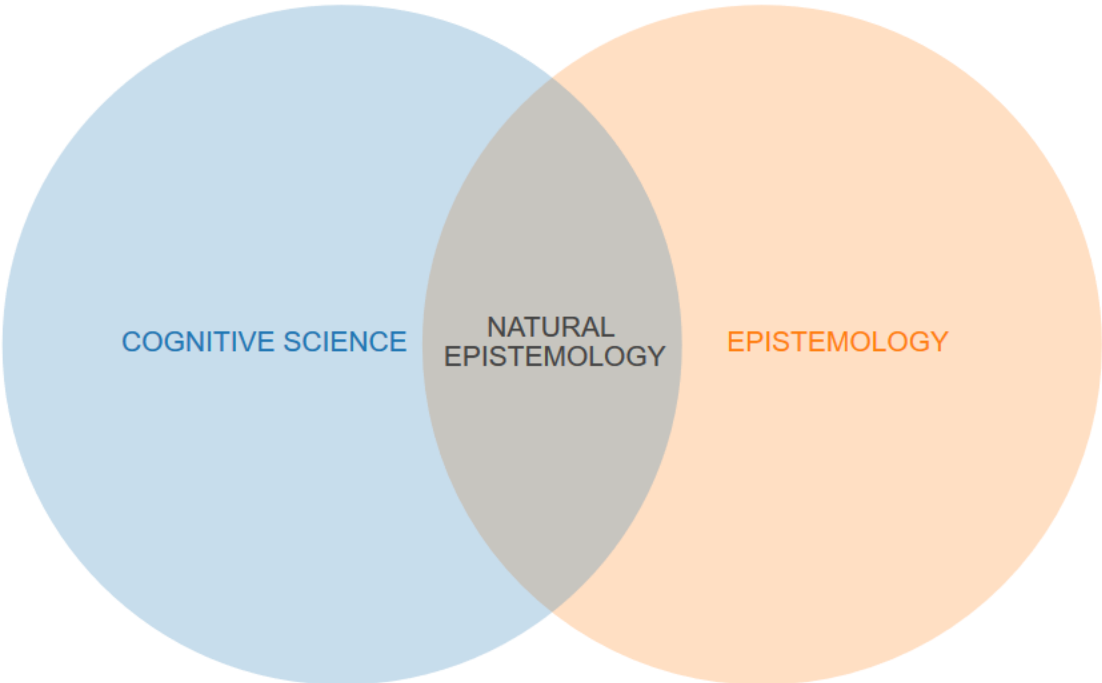

Figure 3. The intersection of cognitive science and philosophy in regards to epistemology is natural epistemology.

1.2 Evolution of Knowing Within Cognitive Science and Evolution of Knowing of Cognitive Science

The exploration of knowing in cognitive science consciously manifests itself in researching subjects and their knowing from a third-person point of view and could therefore be dubbed as third-person epistemology. To show how such research and insights evolved, I will present a few key cognitive phenomena and how cognitive science historically, but also ideas-wise, progressed due to evolving insights on these phenomena. All these phenomena seem to be central to the understanding of the relationship between mind and world. Presenting such a history is important, as it shows researchers' entrapment into the scientific Zeitgeist of their time. A simple, plain illustration to show how such entrapment works, and has worked even before the existence of cognitive science: the brain (and by extension, mind) has always been metaphorically thought of as the most advanced technological piece of the time. Searle (1984) notes how the ancient Greeks thought of it as a catapult, the great polymath and philosopher Gottfried Wilhelm Leibniz as a mill, the Austrian psychoanalyst and neurologist Sigmund Freud as a hydraulic and electro-magnetic system, and the prominent neuroscientist Charles Scott Sherrington as a telegraph system. These metaphors shape research, concepts, theories, general scientific work as well as broader conceptualisation of the world which holds global societal implications. This is why unravelling history and showing how these conjectures come into being and interact is so important for understanding the light in which cognitive science is presented in this thesis. At the end of this section, I will try to summarise what happened in the examples provided through a more abstract, loose framework.

Perception has been, due to its relevance to this thesis as the central phenomenon, an ever-present theme so far. It has been so far presented as a common grounds for the relationship between epistemology and cognitive science (through Berkeley and computer vision), and it will now be used to show the crucial process of cognitive science as natural epistemology. It is for this reason that perception has to be put in the context of cognitive science and its history. Vision, being the most accessible of the senses (Palmer, 1999), especially for experimental and synthetic testing, has consequently one of the biggest research accounts in cognitive science. And with the biggest accounts almost by rule come the biggest problems as well. Historically, the breeding grounds for vision research was what was 50 years ago synonymous to cognitive science – artificial intelligence. It was computer vision that was the source of knowledge about vision, the testable tool for various theories and for finding out what happens between the outside world and visual experience (Vernon, 2005). To understand the state of research on (computer) vision, it is important to understand how mind (and with it, cognition) was thought of in the early years of cognitive science, which was not much different from today.

Mind was seen as a computer that "entails the manipulation of explicit symbolic representations of the state and behaviour of an objective external world" (Vernon, 2005, p. 6). This view or paradigm is called cognitivism, and is still the prevailing one in cognitive science. Computer, algorithmic, information-processing vision was therefore seen as a true duplicate of biological vision (Marr, 1982). The view on mind and perception as symbolically representing the objective external world gives a strong clue on what the relationship between mind and world supposedly is. Considering visual perception in this way raised a number of insurmountable barriers, which became apparent exactly through trying to replicate visual perception with computer vision. The latter was slow, could not perform real time, was generally full of errors, was extremely limited in its scope and had considerable issues performing in domains that were not completely specified (Moravec, 1988). Marr (1982, p. 31–37) established the definition of vision as “a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information”, where “vision alone can deliver an internal description of the shape of a viewed object” (Ibid.). But since, according to cognitivists Hoffman, Singh and Prakash (2015), “having a perceptual experience does not require motor movements”, which is similar to the general cognitivist separation of perception and generation of behaviour (Möller, 1999), the agent possessing such vision would have no idea about what inputs are irrelevant, which means that “all aspects of the visual input have to be considered as potentially relevant for the generation of (arbitrary) actions” (Ibid., p. 169), resulting in very high complexity where context is impossible to depict. The problems machine vision has been having therefore makes perfect sense in light of such conceptual shortcomings. Another, so far unsolved challenge for algorithmic accounts of vision is the gap between the algorithm and experience, which becomes apparent in certain visual phenomena such as constitution of the visual world through saccadic movement, blind spot, change blindness and so on (Blackmore, Brelstaff, Nelson, & Troscianko, 1995). The construction of the visual world and the role of saccades is well investigated, and offers meaningful insight into the problem of cognitivist views on visual perception. A saccade is a fast, consciously undetectable eye movement from one position to another. In humans, up to five saccades occur per second (Hancock, Gareze, Findlay, & Andrews, 2012). A saccadic movement is not smooth, contrary to our visual experience – it jumps around, with gaps in-between. The reasons for these gaps not being perceived are twofold: it is too fast to be detected, and at the same time it is inhibited by top-down visual processing which fills the gaps. Jug, Kolenik, Ofner and Farkaš (2018) argue that top-down visual processing is required to consciously experience the visual world as we do, as otherwise our visuals would be constantly going in and out of experience. This rectangular picture that we experience as if it is transmitted bottom-up before us is an illusion – the saccades, going from position to position, and top-down filling of the gaps construct this stable, whole image. This combination presents a crucial difference between biological visual perception and computer vision. The first constructs the visual world by bottom-up saccades that add to it bit by bit and top-down filling up of the gaps, while the second snaps the scene before the sensors as it is in a certain moment, as a whole. This role of top-down processing and attention-seeking behaviour through saccades towards subjectively-salient parts of the world is unaccounted for in cognitivist accounts of vision; taking this role into account, computer models deal with visual perception exceedingly well (Jug et al., 2018).

The two examples of visual phenomena led to a certain rethinking of positions on the mind-world relationship in cognitive science. The presuppositions about it needed to be re-evaluated in order to give accounts of visual perception a chance at overcoming many obstacles, as well as to accommodate findings from other disciplines in cognitive science. In artificial intelligence, roboticist Rodney Brooks recodified the mind-world relationship in a way that helped make a breakthrough: “Just as there is no central representation there is not even a central system. Each activity producing layer connects perception to action directly. It is only the observer of the Creature who imputes a central representation or central control. The Creature itself has none; it is a collection of competing behaviors. Out of the local chaos of their interactions there emerges, in the eye of an observer, a coherent pattern of behavior” (Brooks, 1991, p. 148–149). Such reconsiderations led to a wider paradigm-shift, which found home in a conglomerated family of enactivism:

> Enactive systems take the emergent paradigm even further. In contradistinction to cognitivism, which involves a view of cognition that requires the representation of a given objective pre-determined world, enaction asserts that cognition is a process whereby the issues that are important for the continued existence of the cognitive entity

> are brought out or enacted: co-determined by the entity as it interacts with the environment in which it is embedded. Thus, nothing is 'pre-given', and hence there is no need for symbolic representations. Instead there is an enactive interpretation: a real-time context-based choosing of relevance. The advantage is that it focusses on the dynamics by which robust interpretation and adaptability arise. (Vernon, 2005, p. 7)

The sensorimotor loop, which was non-existent in cognitivism and among other problems caused the frame problem for perceiving agents, and structural coupling between the agent and the world, which results in personal niche enaction, seem to be a way of looking into more promising ways of potentially solving the aforementioned problems in theories of visual perception (Jug et al., 2018) and epistemological questions in general.

What happened during these paradigm shifts is crucial for conceiving cognitive science as natural epistemology. Researching knowing appears to have a double effect: it changes how cognitive science perceives knowing, but it also changes the knowing of cognitive science itself, changing fundamental epistemological presuppositions. These do not seem to be explicitly investigated as such, as there is very little research that state researchers' presuppositions or, furthermore, investigate them. When research in different phenomena starts hitting a wall, and when this is not a consequence of, e.g., technological limitations, epistemological presuppositions become a likely candidate for investigation. Admittedly, in rare cases, by becoming aware of deeper epistemological issues, which have not yet surfaced up to the point of hitting the wall, research can be expended to embrace them as well, resulting in overcoming the barriers from before. Luhmann has, to a certain degree, observed this pattern and articulated it, albeit in different terms and context: "It is no surprise for a naturalized epistemology to come up against its own self-reference" (Luhmann, 1996, 479). This self-reference has apparently manifested in cognitive science, and following Luhmann, this self-reference has to be included in natural epistemology. This process therefore seems very important in cognitive science and essential if it is to be characterised as natural epistemology. This process can be summarised in a simplified step-by-step format, which loosely exhibits what happens in research, albeit on rare occasions, that may result in breakthroughs:

1. Scientists research a particular cognitive phenomenon with their existing knowledge of cognitive science about cognising. This includes certain epistemological presuppositions.
2. Scientists make great strides, but suddenly hit a wall. The studied phenomenon cannot be explained with current methods.
3. Scientists believe that either more research or sufficient technological advancement relating to their methods will resolve the issue they struggle with in the studied phenomenon.
4. Scientists realise that they have been thinking about the phenomenon in a wrong way. Thinking about it through a different set of epistemological presuppositions opens the doors to novel research on the phenomenon.
5. Scientists find a solution to the previous issue in researching the phenomenon and cause a paradigm shift in how cognitive science sees the relationship between mind and world as well as how it sees mind and world itself. As epistemological insights in cognitive science cause a need for epistemology of cognitive science to change, the latter in turn feedbacks and changes epistemological insights in cognitive science.

This step-by-step *concept of the knowledge feedback loop of natural epistemology* (COKFLONE[2]) is an ideally and loosely described scenario. However, e.g., the example regarding vision and how cognitivist paradigm caused the enactive paradigm to take shape can be fit into this concept. This happens on very rare occasions, but these occasions are of the outmost importance to the field. COKFLONE that forms between knowing within cognitive science and knowing of cognitive science can be seen in Figure 4. The roles of "knowing within cognitive science" and "knowing of cognitive science" interestingly point to two ways of thinking in scientific research as characterised by Kuhn (1959) as "essential tension". "Essential tension" plays out as a contrast between the so-called convergent thinking and divergent thinking. Convergent thinking is what scientists commonly do in their regular work, where the "scientist is not an innovator but a solver of puzzles, and the puzzles upon which he concentrates are just those which he believes can be both stated and solved within the existing scientific tradition" (Kuhn, 1959, p. 234). Convergent thinking is "neither intended nor likely to produce fundamental discoveries or revolutionary changes in scientific theory" (Ibid., p. 233). According to COKFLONE, "knowing within cognitive science" represents Kuhn's concept of convergent thinking. Divergent thinking, however, is when "the scientist must […] rearrange the intellectual and manipulative equipment he has previously relied upon, discarding some elements of his prior belief" (Ibid., p. 226). "Prior belief" that Kuhn notes is suspiciously similar to epistemological presuppositions, held by scientists or paradigms in cognitive science. The landscape of "knowing of cognitive science" is therefore strikingly associated with Kuhn's concept of divergent thinking. Convergent and divergent thinking follow COKFLONE well enough as well, where, to solve problems that cannot be solved with convergent thinking, divergence is needed.

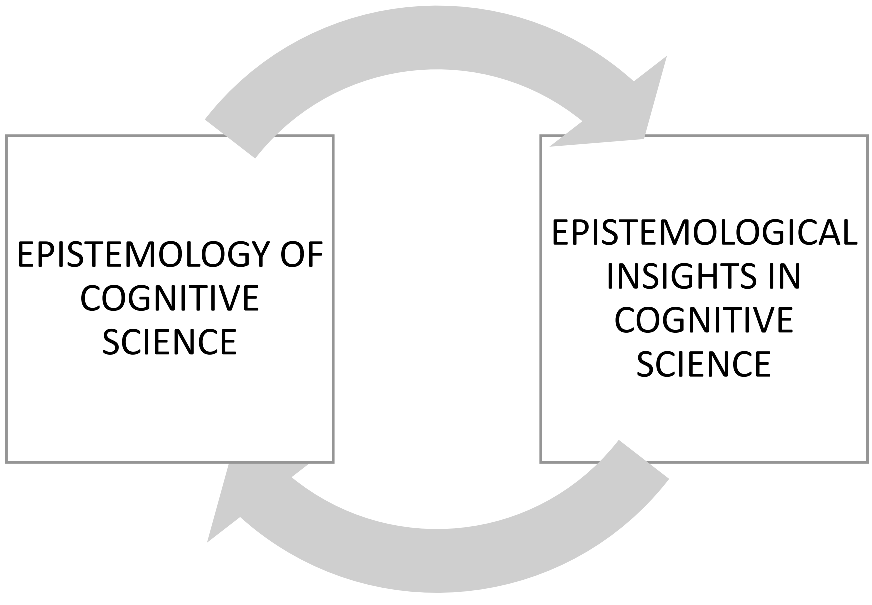

Figure 4. The loop between epistemological insights in cognitive science and epistemology of cognitive science (COKFLONE). The first reflects knowing within cognitive science, the second knowing of cognitive science.

[2] For easier tracking of abbreviations and what they mean, see Appendix A where all the used abbreviations are listed.

Loops that COKFLONE embodies are characteristic of the cybernetic way of thinking that was prominent in the 1940s and 1950s. It is this cybernetic legacy that combined with extreme positions in researching epistemology of researching systems (such as cognitive science) and spawned the field of second-order cybernetics, which shifted the aforementioned third-person epistemology to first-person epistemology – applying the investigations of knowing of living beings to ourselves, to living beings doing the investigations. Von Foerster, the father of second-order cybernetics, delineates first-order and second-order cybernetics: "[…] first-order cybernetics is the cybernetics of observed systems, while second-order cybernetics is the cybernetics of observing systems" (von Foerster, 1995, p. 11). More specifically, second-order cybernetics is "the study of the organization in autopoietic machines, which are able to build their own components, and possess cognizance which allows them to observe and control other machines" (Mancilla, 2013, p. 73). The observer, too, is a cybernetic system, subject to the same processes as the observed, external, "independent" systems. Natural epistemology connects and, in a way, gives a framework to think about how these positions, from cognitive science as third-person epistemology to second-order cybernetics, are connected, and cautions observers to be aware of these positions and circularity between them. By being aware of such implications for research, the researcher has to explicate her position on these issues.

1.3 Knowing of a Cognitive Scientist

This section comes from a place of awareness about my position in the catbird seat in regards to all aspects of my scientific work. Therefore, I am writing this as a kind of *caveat emptor*, a disclaimer. Bateson (1979, p. 93) says that "epistemology is always and inevitably personal." If the researcher were to possess such a stance, it would be necessary for her to acknowledge it. As Bateson (Ibid.) continues: "The point of the probe is always in the heart of the explorer: What is my answer to the question of the nature of knowing?" This echoes the prevalent, or rather the main, sentiment of second-order cyberneticists and, by extension, radical constructivists. The universality of Bateson's words can be found in numerous different phrasings by many of his contemporaries. Von Foerster (2003, p. 289) credits cybernetics for the emergence of this viewpoint: "What is new [in cybernetics] is the profound insight that a brain is required to write a theory of a brain. From this follows that a theory of the brain, that has any aspirations for completeness, has to account for the writing of this theory. And even more fascinating, the writer of this theory has to account for her or himself." Francisco Varela, Evan Thompson and Eleanor Rosch dedicated a comprehensive section in their seminal work *The Embodied Mind: Cognitive Science and Human Experience* (1991/2016) to this issue alone. Titled "Cognitive Science within the Circle" (Ibid., p. 9), Varela et al. say:

> We began this chapter with a reflection on the fundamental circularity in scientific method that would be noted by a philosophically inclined cognitive scientist. […] we cannot avoid as a matter of consistency the logical implication that […] any […] scientific description, either of biological or mental phenomena, must itself be a product of the structure of our own cognitive system. […] Furthermore, the act of reflection that tells us this does not come from nowhere; we find ourselves performing that act of reflection out of a given background (in the Heideggerian sense) of biological, social, and cultural beliefs and practices. […] But then yet again, our very postulation of such a background is something that we are doing: we are here, living embodied beings, sitting and thinking of this entire scheme, including what we call a background. (Varela, Thompson & Rosch, 1991, p. 9–12)

Varela, Thompson and Rosch capture their endeavour in the bottom figure:

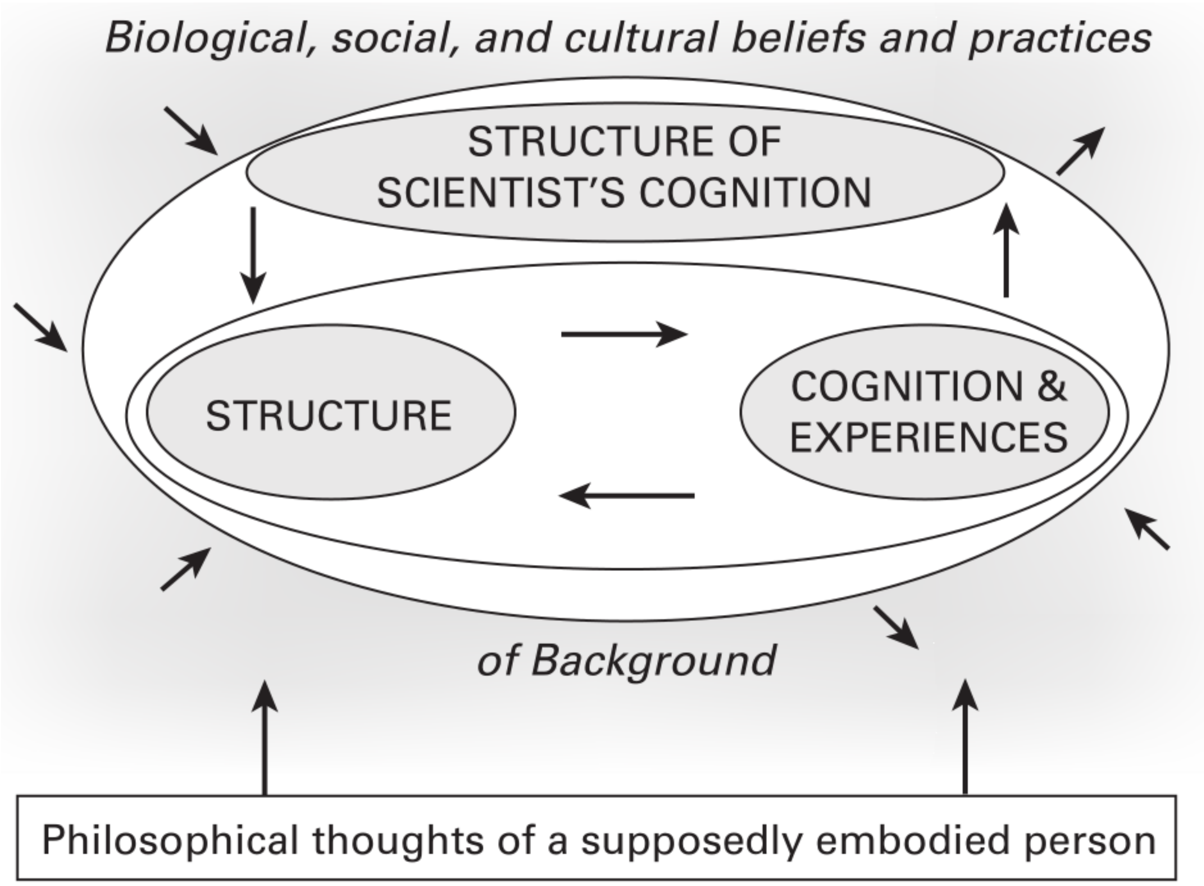

Figure 5. Observing and its interdependency on biological, social and cultural factors (from Varela, Thompson, & Roesche, 1991, p. 11).

There is a possibility that personal biases creep into an otherwise by-the-book research, even when strictly following the scientific method. However, I will not be consciously letting my personal epistemological positions inform my method and research; I hold that valid, legitimate research results are possible by being aware of one's position and taking the necessary steps to prevent biasing. What may occur, though, is influence in the opposite direction – my personal epistemological position and biases may change due to the research results, thus demonstrating the cycle of research and the loop of knowing in natural epistemology.

The thesis is also intentionally written in first person singular and in active voice when appropriate, as opposed to more commonly used third person, plural and passive voice. This is to further acknowledge my stance as described in this section.

Varela et al., even after a rigorous reflection of a researcher's role in her research, admit that such a reflection can only go so far and that it is important to return to existing research at hand, saying that "rather than adding layers of continued abstraction, we should go back where we started, to the concreteness" (Varela et al, 1991/2016).

Starting next chapter, this endeavour of exploring cognitive science as natural epistemology will be becoming more and more precise, more and more focused on concrete phenomena and methods. One of the most interesting areas of research for epistemological investigations is revealing itself to be perception, which is why a large part of the thesis has already been dedicated to it. Looking at perception through an evolutionary lens may unveil what the role of perception is in the relationship between mind and world. This is why the next chapter will explore perception through an evolutionary perspective.

2 Evolutionary Perspective on Perception

In this chapter, the prevailing views on perception – which is an essential part of the relationship between mind and world (Kvanvig, 2003) – will be presented from an evolutionary perspective. Through different evolutionary and especially biological accounts, the problems with the prevailing views on perception will be brought to light, followed by a discussion on how the puzzling questions that arise from the problems can be studied.

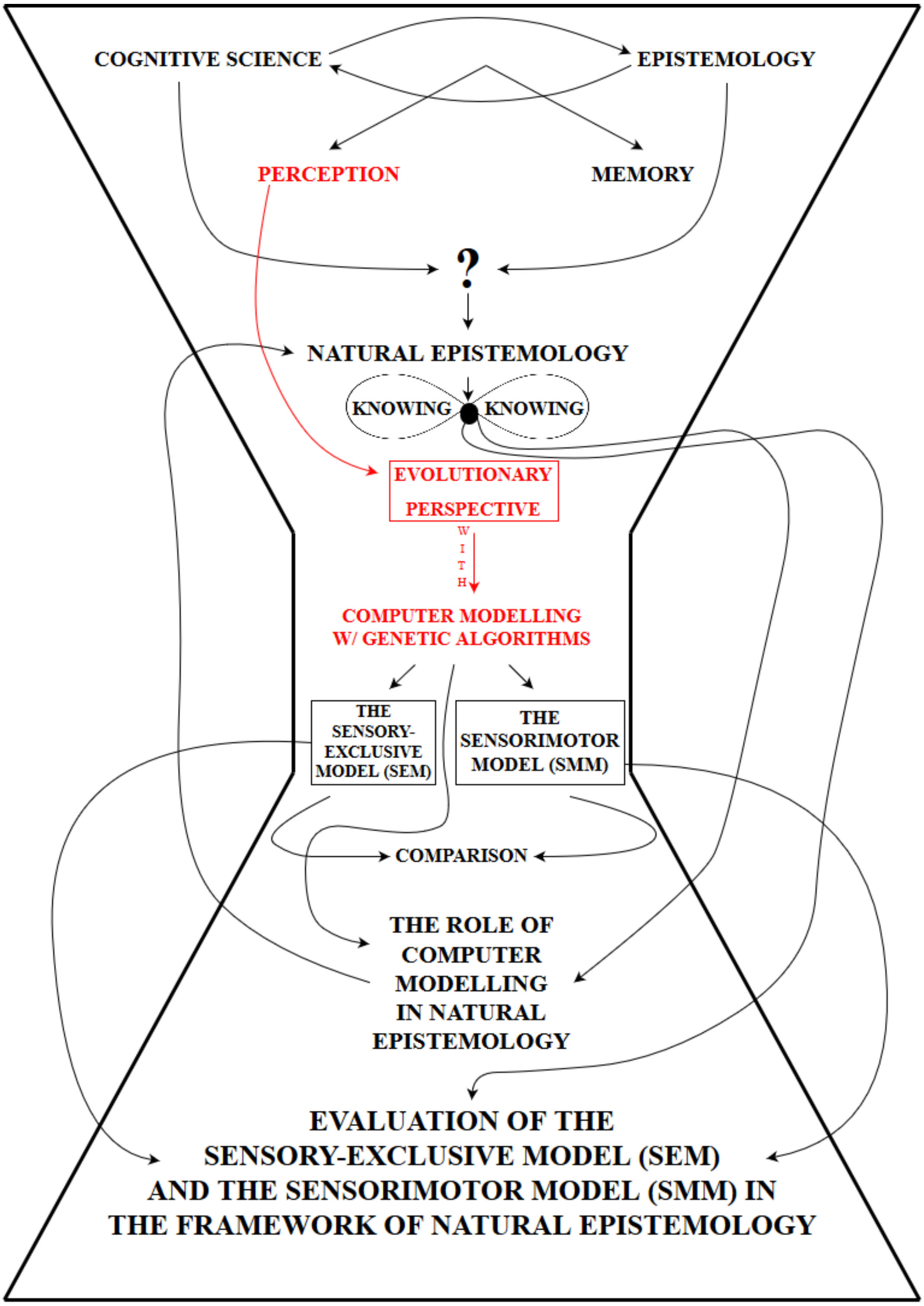

Figure 6. The position in the narrative schematic of the thesis, marked in red.

2.1 Perception and How to Perceive It: Prevalent Views on Perception from an Evolutionary Perspective and Its Problems

Lehar (2003, p. 375) offers a succinct, yet rich summary of how various sciences, be it biology, psychology, neuroscience or artificial intelligence, view perception, how it functions and what it is for: "The primary function of perception [is] that of generating a fully spatial virtual-reality replica of the external world in an internal representation" (Lehar, 2003, p. 375). Such perception is supposed to have evolved in organisms "to match the structure of the world and the coding scheme provided by nature" (Knill & Richards, 1996, p. 6). This position on perception and its evolution is mostly taken for granted where science is concerned. Both college textbooks (Friedenberg & Silverman, 2016; Palmer, 1999) and major works by the most prominent scientific authorities (Craik, 1943; Fodor, 1985; Marr, 1982) generally take this position to be true. Even contemporary Bayesian theories, which are often presented as alternatives to established views in their conceptualisation of various cognitive phenomena (Bowers & Davis, 2012), and thus often disagree in stances, claim that perceptual estimates that are closer to "truth"[3] (which is, according to the established views, the outside, objective world), are more useful than those that are further away from it (Geisler & Diehl, 2003). This is almost the same stance as "visual perception is useful only if it is reasonably accurate," articulated by Palmer (1999, p. 6). Generally, cognitive science is no exception to this perspective – that "we have what is called veridical perception (from the Latin veridicus meaning to say truthfully): perception that is consistent with the actual state of affairs in the environment" (Palmer, 1999, p. 6) – but what is more, cognitivism, the prevalent view on the mind in cognitive science, mostly holds this position at its core (Bermúdez, 2014; Lungarella, Lida, Bongard, & Pfeifer, 2007; Palmer, 1999; Pecher & Zwaan, 2006). However, it is Palmer himself, along with many of his contemporaries, who claims that perception evolved exclusively "to aid in the survival and successful reproduction of organisms" (Palmer, 1999, p. 6). It is therefore not entirely clear whether perception's (and cognition's) function is to faithfully mirror the outside world, meaning that "by and large, what you see is what you get" (Palmer, 1999, p. 6), or to serve evolutionary needs of the organism. Peter J. Graham, another evolutionary biologist, argues that "evolution does not care about veridicality" and that "nature does not select for truth and accuracy as such" (Graham, 2014, p. 19). However, he continues: "I think perceptual states contribute to fitness by accurately representing the environment" (Graham, 2014, p. 19). His "think" reflects a dangerous habit, prevalent in his and his colleagues' discourse. This intuition that perception is "a perfectly clear window onto reality" (Palmer, 1999, p. 6) that necessarily serves evolutionary fitness hints at a paradox when coupled with their own claims that "the relationship between a living organism and some element of its environment is determined primarily by how that element is 'perceived' by the living organism relative to its needs" (Falk, 2018, p. 99) which has to be understood as evolutionary "adaptations, […] a compromise between different needs of an organism" (König, 2001, p. 10395).

It seems that the function of perception cannot be that of making an accurate picture of the outside world through passive intake of information from the environment into the mind. Even when considering the possibility of a case where an organism's functioning leads to its

[3] I am intentionally using the same word that is used by the cited scientists to avoid confusion and unnecessary paradigm digressions.

perceptions presenting the objective world in an accurate representation, such convergent phenomenon makes little sense and even less chance considering that organisms perceive according to their internal, unique needs (von Uexküll, 1982). How these should be considered is best described by von Uexküll (1982), who paints a door to unique worlds of other living beings by inviting his readers to

> first blow, in fancy, a soap bubble around each creature to represent its own world, filled with the perceptions which it alone knows. When we ourselves then step into one of these bubbles, the familiar meadow is transformed. Many of its colorful features disappear, others no longer belong together but appear in new relationships. A new world comes into being. Through the bubble we see the world of the burrowing worm, of the butterfly, or of the field mouse; the world as it appears to the animals themselves, not as it appears to us. This we may call the phenomenal world or the self-world of the animal. (von Uexküll, 1957, p. 5)

There is an overwhelming number of studies on animals that show that they, indeed, perceive uniquely, though in less colourful words than von Uexküll's. Hoffman et al. (2015) present two such examples:

> Dragonflies, for instance, have aquatic larvae and must find water to lay their eggs. Dragonfly vision has a simple trick to find water: Find horizontally polarized light reflections […]. Water strongly reflects horizontally polarized light, so this trick often guides successful oviposition. Unfortunately for the dragonfly, oil slicks and shiny tombstones also reflect such light, sometimes more strongly than water. Dragonflies are fooled by such slicks and tombstones to lay eggs where they cannot survive. In the niche where dragonflies evolved, their perceptual strategy normally works, but where that niche has been disturbed by H. sapiens with oil slicks and tombstones, the same strategy can be fatal.
>
> Male jewel beetles fly about looking for the glossy, dimpled, and brown wing-casings of females. When males of H. sapiens began tossing out empty beer bottles that were glossy, dimpled, and just the right shade of brown, the male beetles swarmed the bottles and ignored the females, nearly causing the extinction of the species […]. The beetles' perceptions relied not on veridical information but rather on heuristics that worked in the niche where they evolved. (Hoffman et al., 2015, p. 1481)

The two examples show how diverse perceptions can be. The first distinction between perceptions, important for this thesis, is the difference between *veridical* and *non-veridical* perceptions. Hoffman (2018, p. 731) defines *veridical perceptions* as "perceptions that accurately describe those aspects of the environment that are crucial to survival and reproductive fitness". *Non-veridical perceptions* are the opposite of veridical perceptions. In the two examples, it is non-veridical perception that heuristically guides the animals to evolutionary successes. The experience they have of the world is not its faithful representation. Careful reading of Hoffman reveals that non-veridical perception can be of two kinds, *isomorphic* or *non-isomorphic*, while veridical perception can only be *isomorphic*. *Isomorphism* "implies a structure-preserving relation between the physical-causal make-up of the system and the formal structure of the computational model supposedly instantiated by the system" (Haselager, de Groot, & van Rappard, 2003, p. 7). In regards to the dragonfly and the male beetle in Hoffman et al. (2015), their perceptions still preserve the structures of the outside world, namely the reflections of the surfaces which are isomorphically represented through their perceptions.

The puzzle of perceptual veridicality need not be only discussed in terms of minute comparing of different perceptions from biological sources. In the following section, possible methodology that can give insight into evolutionary perspective on veridical and non-veridical perception, and thus insight into the epistemological question of the relationship between mind and world, will be discussed.

2.2 Perception and How to Evolve It: Methodology for Studying Perception from an Evolutionary Perspective

When looking at it closely, the dispute on the role of perception and how that role is manifested through perception and its functioning is fundamentally an epistemological question. Perception as one of the sources of knowledge ultimately defines and describes the relationship between mind and world, while at the same time defining how organisms cognise and know. However, the mystery of perception as an epistemological issue can hardly be researched analytically, i.e. through observation and experimentation, and therefore has to be subjected to synthetic approaches, i.e. by making computer and robot models (Mirolli & Parisi, 2011). Synthetic approach to (cognitive) science has been called upon by various field authorities, such as Harvey (2000), Froese and Ziemke (2009), to tackle the problem of opposing theories of epistemological nature. These researchers have already given research in AI "the rather privileged position of being able to help resolve theoretical disputes which have plagued the Western philosophical tradition for decades if not centuries" (Froese, 2007, p. 6) through understanding by building (Pfeifer & Scheier, 1999). Since the problem to be tackled pertains to evolution, one synthetic method can offer tools to study perception under the mechanism of natural selection – genetic algorithm.

Froese's sentiments, which are pretty radical in the statement, are worth exploring, and a large part of this thesis will be concerned with the relationship between AI as methodology and philosophically-inclined theories. To study the influence of evolution on perceptual veridicality, appropriate computing methods have to be selected. Genetic algorithms have been used to investigate evolutionary theories (Hamblin, 2013). They are, therefore, a good candidate for such research. Computer modelling and genetic algorithms as methods for natural epistemology are discussed at length from chapter 4 onwards.

Genetic algorithms may therefore provide a way to study the performance of artificial organisms with either veridical or non-veridical perception and compare their faring under the pressures of natural selection. One of the few if not the only research that studies non-veridical perception with genetic algorithms is Hoffman's, Prakash's and Singh's work (2015) on Hoffman's interface theory of perception (ITP). Specifically, Hoffman et al. study whether perception is isomorphic or non-isomorphic to the outside world.

2.2.1 Evolving Perceptions: A Natural Epistemology Approach

Despite the lack of computer models where perceptual veridicality is studied from the evolutionary perspective, there are attempts of modelling the influence of natural selection on perceptual veridicality. Hoffman et al. (2015) use genetic algorithms to test Hoffman's interface theory of perception. The interface theory of perception claims that perception is not attuned to “truth” and does not convey the outside, objective world as it is, but is rather attuned to fitness, meaning that organisms perceive and experience the world in a way that is important to them evolutionary, for survival. Perception therefore works as an interface for organisms to access the world. This interface is analogue to the computer screen, meaning that an icon that is a particular colour and the fact that we can delete it by moving it and dropping it into the Recycle Bin does not represent the true nature of the computer, the transistors, the electrical current, the circuits, etc. According to Hoffman et al., their computer model shows that under the pressures of natural selection, the organisms evolve non-veridical, specifically non-isomorphic perception, attuned to fitness rather than “truth”.

This phenomenon and Hoffman et al.’s work present an interesting case study that can be put into the framework of natural epistemology and investigated, and this can happen on two levels: 1) their claim can be investigated through other paradigms in cognitive science, therefore questioning the authors’ presuppositions and testing the claims under different ones, 2) the methodology’s usefulness and implications can be evaluated in the framework of natural epistemology. The latter provokes several questions: What is the connection between computer modelling and broader epistemological presuppositions, which serve as a foundation for all paradigms in cognitive science? Can computer modelling reveal anything about them and about their influence on the studied phenomena? Can computer modelling, and natural epistemology in general, help philosophy in answering its epistemological questions and point into the right direction in regards to the ideas about the mind and cognition? How can computer modelling serve to advance cognitive science’s views on the mind? Is computer modelling with genetic algorithms a (good) methodological example of and for natural epistemology?

The goal of these questions is not centred on the concrete results from the modelling as such – my goal is to conduct a research with computer modelling that tackles epistemological questions and evaluate its usefulness for the natural epistemology endeavour. To achieve that, I reproduce a genetic algorithm model by Hoffman, Prakash and Singh (2015) and rework it into a new, yet comparable model with its own presuppositions for comparison and consequent analysis of synthetic methodology. In the analysis, I will be particularly interested in how (well) it fits into the framework of natural epistemology. Before presenting the models, the research goals and the methods are presented.

3 Research Questions and Goals

This research has two goals (RG). The first one is general and permeates the entire thesis, while the second one is necessary to achieve the first one, as it presents an example of the endeavours of the first research goal:

RG1) Analyse and critically assess the value and role of computer modelling with genetic algorithms in understanding epistemological presuppositions in models of cognition, as well as determine the value of computer modelling with genetic algorithms for studying epistemological questions in the framework of natural epistemology. Demonstrate by making a computer model with genetic algorithms and evaluating it.

To fully grasp and legitimately work on this research goal, two computer models will be constructed. The main purpose is to learn from them and see first-hand what utilising computer modelling as a methodology for epistemological inquiry would look like:

RG2) Use synthetic methodology, computer modelling with genetic algorithms, to research the evolution of perceptual veridicality under the influence of natural selection. I am interested in what kind of perceptions have higher survival value for organisms, isomorphic or non-isomorphic perceptions. The computer modelling will be based on the work by Mitchell (1999) and Hoffman et al. (2015). RG2 consists of three research questions (RQ):

RQ1) Does the reproduction of the model by Hoffman et al. (2015) produce the same results as the original model does?

The genetic algorithm model by Hoffman et al. (2015), which is based on the Robot Robby genetic algorithm by Mitchell (1999), shows that for their artificial organisms, non-isomorphic perception presents higher survival value.

The hypothesis is that the reproduction will be successful, thus yielding the same result as Hoffman et al.'s model produces.

RQ2) Which presuppositions does the model by Hoffman et al. (2015) hold?

RQ2 is necessary to follow the framework of natural epistemology and dig deep into how presuppositions influence synthetic research in epistemology. The reason behind reproducing the model of Hoffman et al. was largely to answer this research question. In-depth understanding of it was needed to discern the presuppositions of the authors. The authors may or may not have been aware of these presuppositions.

The hypothesis is that the presuppositions in the model by Hoffman et al. are analogous to certain cognitivist presuppositions.

RQ3) How can the computer model by Hoffman et al. (2015) be modified so that certain cognitivist ideas about cognition, which manifest themselves in certain presuppositions, are changed with certain enactivist ideas and presuppositions? How does this change affect the results?

To answer RQ3, a more comprehensive analysis of enactivist presuppositions in relation to computer modelling is needed. It is especially important to discuss which enactivist ideas are compatible with computation and can be modelled, as well as evaluate which enactivist ideas are sensible to be included given the nature of Hoffman et al.'s model.

The hypothesis is that the results of the model with enactivist presuppositions will be generally the same as the results of Hoffman et al.'s model. I suppose there will be certain problems in reconciliation between computer modelling and fundamental enactivist ideas.

4 Methods and Procedures

To achieve RG1, a long-form critical analysis is needed. The critical analysis is inseparably embedded into the whole thesis as such. The articulation of cognitive science as natural epistemology, the COKFLONE, the discussion on evolutionary perspective on perception and its issues, the design and implementation of computer models, the critical analysis of computer modelling and models themselves, each part of the thesis directly serves and has a place in achieving RG1. This should be apparent with all the different parts of the thesis building on one another to tell a larger story by examining its subject matters closer and closer, on lower and lower levels, from cognitive science as a particularly imagined discipline to a very specific computer model and back to cognitive science as natural epistemology.

To achieve RG2, synthetic methodology will be used. Mirolli and Parisi (2011) advocate for the use of the synthetic approach to science, claiming that computer models can be used to explain “real phenomena” (Ibid., p. 298), as models are generally seen as one possible explanation of “reality”. Synthetic methodology’s modus operandi is largely that of understanding by designing and building (Pfeifer & Scheier, 1999). Specifically, genetic algorithms were used for this research.

A genetic algorithm (Mitchell, 1999) is an optimisation heuristic, which works well for problems where the optimal solution is not known or is too computationally demanding. A genetic algorithm (GA) is used to yield a satisfactory solution. Mostly, GAs have been used to solve practical problems, as it has long been ignored or snubbed by computational biologists. Mitchell argues that using simulations has benefits in biologically-inclined areas of research:

> A computer program can simulate the evolution of populations of organisms over millions of simulated generations, and such simulations can potentially be used to test theories about the biggest open questions in evolution. Simulation experiments can do what traditional methods typically cannot: experiments can be controlled, they can be repeated to see how the modification of certain parameters changes the behavior of the simulation, and they can be run for many simulated generations. (Mitchell, 1999, pp. 65–66)

The use of GAs in cognitive science has been tentative as well. GAs are mostly used as an optimisation technique (Whitley, 1994), which is (also) why they most often than not rely on arbitrary and trivial parameters (Eiben & Smith, 2003). This makes them removed from biology, but GAs can still be useful when having fixed expectations in a limited set of phenomena. The popularisation of artificial neural networks for researching cognitive phenomena can be seen as a consequence of certain limitations of GAs. Nonetheless, GAs enshrine the mechanism of natural selection, which is very useful for researching what passes the test of adaptation and survival, especially when that includes comparing different paradigms and theories.

GAs represent all the general attributes of natural selection, especially gene recombination with mutation. In GAs, individual organisms possess DNA, which is transcribed as one or more chromosomes. The latter are made of several genes, which can be strings of bits, integers, characters, functions, etc. They are expressed in an organism’s phenotype, having an important influence on the organism’s behaviour.

A population is produced in the GA by running a procedure that yields organisms with different DNAs. Of the utmost importance is the fitness function, which represents the environmental pressures on the organisms and determines which organisms cope best in the environment. The organisms with the highest fitness scores usually have a higher probability of reproducing.

Genetic crossover with mutation ensures that the population adapts to its niche, evolving progressively better individuals with higher fitness scores. Usually, best-performing organisms are selected and paired together. They create an offspring, which takes the first half of the chromosome from one organism and the second half of the chromosome from the second organism. There is a certain probability for mutation of the newly created DNA's genes in this process. A certain probability for organisms with low fitness scores reproducing is specified as well, usually to avoid reaching local maxima too quickly. Such probabilities are set arbitrarily, fine-tuned to ensure the best results.

Abstractly, a GA follows a simple procedure:

1. Generate a random population of organisms which represent solutions to a specified problem.
2. Evaluate the solutions in the population using a fitness function.
3. Select solutions from the population proportionate to their fitness score, and perform crossover and mutation to generate a new population.
4. GoTo step 2.

The models that I implement for the thesis will be programmed in Python. For transparency and reproducibility reasons, the code will be uploaded to a commonly-used web-based hosting service Github, which uses Git for version control. The links will be attached to the appropriate sections of the thesis. The code will be uploaded under the MIT License (see Appendix B). In my opinion, the MIT License is the most appropriate license for scientific projects that use code, as it encourages openness and availability for dissemination of knowledge to anyone and for anyone. The MIT License is short and concrete, allowing anyone to do anything with the code as long as they use the MIT License as well, therefore spreading open source and open science principles.

5 The Interface Theory of Perception (ITP)

In this chapter, Hoffman's interface theory of perception (ITP) will be presented and related to the epistemological questions this thesis has explored so far. The GA, used to study ITP, will be studied to reproduce the GA model by Hoffman et al. (2015). Afterwards, the process of reproduction of the GA model will be presented. The presented reproduced model will serve for my analysis of its presuppositions, relevant for this endeavour in relation to cognitive science as natural epistemology.

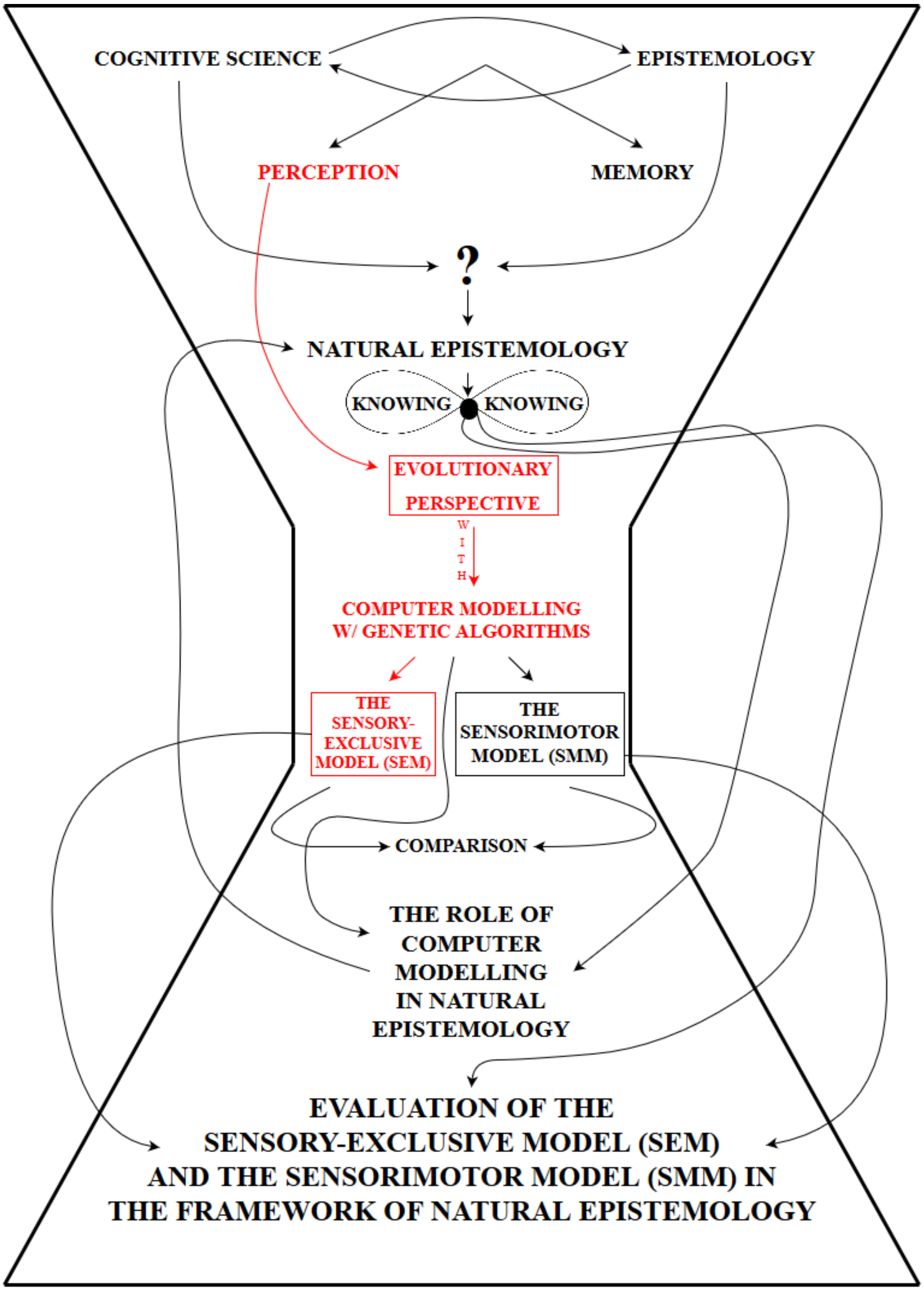

Figure 7. The position in the narrative schematic of the thesis, marked in red.

The interface theory of perception (Hoffman et al., 2015) tries to empirically answer two fundamental questions: "First, is the vocabulary of our perceptions isomorphic to aspects of objective reality so that our language of perceptions could, in principle, describe the objective truth? Second, if so, do our perceptual systems, using that vocabulary, in fact succeed in describing the true state of affairs in the world?" (Hoffman et al., 2015, p. 1482) The main idea of ITP is that the "perceptions of an organism are a user interface between that organism and the objective world" (Hoffman, 2009, p. 7). Hoffman et al. argue that this interface does not show the world as it is. The interface is not attuned to "truth", as that would not be evolutionary beneficial to the organism – if organisms were to have perceptions that would be isomorphic to the outside world, they would quickly die out. The perceptual interface therefore shows the world in a way that it benefits the organism. Its perceptions reflect its internal biological needs and work as a pragmatic guide to navigating the world. Hoffman et al. (2015) use an analogy of a computer to describe the perceptual interface: "A desktop interface makes it easy to use the computer. To delete or copy files, for instance, one simply needs to drag icons around on the desktop. But a desktop interface does not make it easy to know the true structure of a computer — its transistors, circuits, voltages, magnetic fields, firmware and software. Indeed, it's in part by hiding this complex structure that the desktop makes it easier to use the computer" (Ibid, p. 28).

Hoffman et al. (2015) are interested in demonstrating that this interface shows the world non-isomorphically to the reality out there. The most interesting part of the theory is therefore the empirical work, which seems to be a thoroughly novel achievement. The authors use a number of mathematical methods to gauge the theory, including evolutionary games, GAs and Bayesian statistics. They continue to look into other, similar mathematical tools that would make the theory more legitimate (Fields, Hoffman, Prakash, & Prentner, 2017). In the next section, the reproduced GA model will be presented.

5.1 Hoffman et al.'s Sensory-exclusive Model (SEM) and Its Reproduction

This section will introduce the sensory-exclusive model (SEM) as used by Hoffman et al. (2015) to empirically validate ITP. SEM will be described, which will include the description of a model called "Robby, the Soda-Can-Collecting Robot" (Mitchell, 2009, p. 130), designed by Mitchell (2009), and afterwards, the process of reproduction will be presented, which will include the results.

5.1.1 Description of the Sensory-exclusive Model (SEM)

The GA model by Hoffman and colleagues, which will be dubbed as the sensory-exclusive model (SEM), is based on a more commonly known computer model by Mitchell from her book *Complexity: A Guided Tour* (2009), an introductory book on complex systems, where she demonstrates a number of complexity concepts with mathematical and computer models. One of them is "Robby, the Soda-Can-Collecting Robot" (Mitchell, 2009, p. 130). Over many generations, Robby evolves an efficient foraging strategy (the collection of moves associated with individual scenarios) for collecting soda cans on a 10×10 grid of squares, surrounded by a wall, represented by a perimeter of squares. For each square that is not part of the wall, there is a certain probability that it contains a soda can. The visual representation of this scenario can be seen in Figure 8.

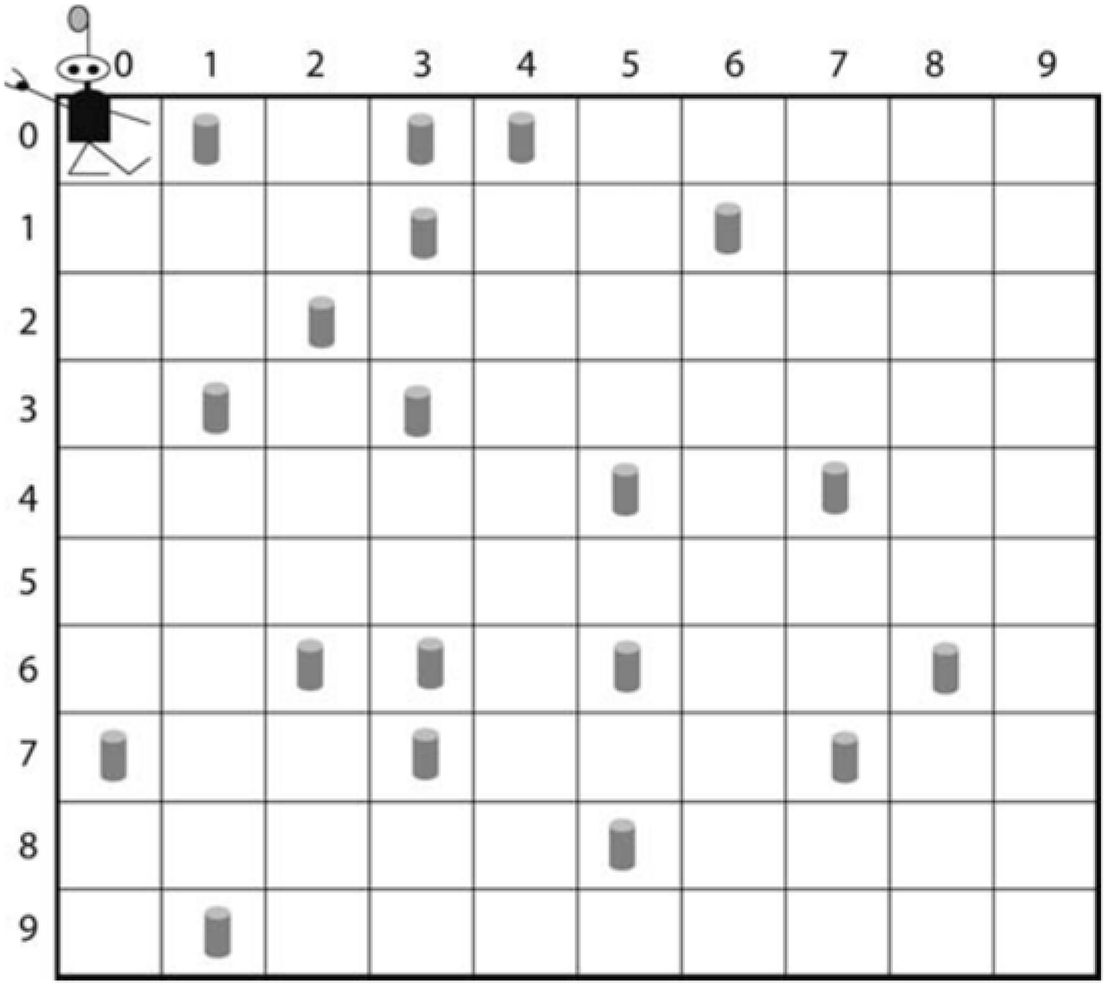

Figure 8. The visual representation of the model (from Mitchell, 2009, p. 131).

Mitchell describes what happens in this model:

> For each cleaning session, Robby can perform exactly 200 actions. Each action consists of one of the following seven choices: move to the north, move to the south, move to the east, move to the west, choose a random direction to move in, stay put[4], or bend down to pick up a can. Each action may generate a reward or a punishment. If Robby is in the same site as a can and picks it up, he gets a reward of ten points. However, if he bends down to pick up a can in a site where there is no can, he is fined one point. If he crashes into a wall, he is fined five points and bounces back into the current site. (Mitchell, 2009, p. 131)

Hoffman et al. (2015) offer a more technical description of Mitchell's model, which they later use and modify for their own ideas:

> [T]he world, call it *W*, that Robby inhabits can be represented as a 12 × 12 grid of squares. We denote the state of square (i, j) by $W(i, j)$ and stipulate that its value is 0 if the square has no cans, 1 if it has one can, and 2 if it is a wall. Because the wall is fixed,

[4] meaning "stay on current position"

> and the state of each square of the inner 10 × 10 grid is either 0 or 1, the possible states of W are $2^{10\times10} = 2^{100}$.
>
> Robby can only see the state of the square he occupies and of the four immediately adjacent states. For instance, if Robby is at location (i, j) then he sees the world states $(W(i,j), W(i,j+1), W(i,j-1), W(i+1,j), W(i-1,j))$. Because there are at most three states at each of these five locations, the space of Robby's possible perceptions, call it X, is no larger than $3^5 = 243$ […]. Robby does not know which square (i, j) he is in, or even that he is in a 12 × 12 grid; he only sees the states of the squares, but not the locations or structure of the squares.
>
> […] Robby has a set, call it G, of seven primitive actions he can take: stay where he is, pick up a can, step north, step south, step east, step west, or step randomly. What must be learned […] is a foraging strategy that specifies which of the seven actions in G to take for each possible perception of the roughly 240 in X […]. The set of possible foraging strategies is thus approximately of cardinality $7^{243} \approx 2.3 \times 10^{205}$ […].
>
> […] Robby gets 10 points for each can he picks up, but loses 1 point each time he tries to pick up a can where there is none, and loses 5 points each time he tries to walk into a wall.
>
> There are roughly 240 'genes' that the genetic algorithm evolves, each having seven possible values, corresponding to the seven actions that can be taken in response to each potential perceptual state. Mitchell starts the genetic algorithm with an initial generation of 200 robots, each having randomly chosen values for each gene. Each robot is forced to forage through 100 randomly chosen worlds, taking 200 actions in each world. The fitness of a robot is the average number of points it collects over its 100 foraging runs. The fitter robots are preferentially chosen to be parents for the next generation. The genes for two parents are randomly split into two parts, and the parts swapped to create two new genomes. A small amount of mutation is applied. In this way a new generation of 200 robots is created, and their fitness again measured by their success at foraging. This process is repeated for 1000 generations. (Hoffman et al., 2015, pp. 1487–1488)

This GA starts out with Robbies with very bad strategies, bumping into walls and picking up cans from empty squares. Over time, Robbies in later generations relish good strategies, picking up a good number of cans and hardly receiving any negative scores.

Hoffman et al. (2015) and Mark (2013) change Mitchell's model in a number of ways to be able to study ITP. Mitchell's Robby only evolves its foraging strategy, while Hoffman's Robby also evolves what the authors call perceptual strategy, which represents a matrix with information for mapping the external world to colours, namely the number of soda cans to a particular colour. Mitchell's Robby sees the world as it is – this is what Hoffman dubs as realist perception. Robby's perceptions therefore have to be changed so that it will be able to either see the world isomorphically or non-isomorphically. This means that the DNA for its perceptual strategy has to be added. Hoffman et al. dub their model as a showcase for "coevolution of foraging and perceptual strategies" (Hoffman et al., 2015, p. 1488). These are the concrete changes made to the algorithm:

> [E]ach square [has up] to […] 10 cans, and […] the following payoff function: (0,1,3,6,9,10,9,6,3,1,0). For instance, a robot gets 6 points for grabbing the cans in a square having 3 or 7 cans, and 0 points for a square having 0 or 10 cans. However, each robot cannot see the exact number of cans in each square, but instead sees just two colors, red and green. Each robot thus has a perceptual strategy, namely a mapping that assigns the percept red or green to each of the 11 possible numbers of cans. Perhaps, for instance, it sees red if a square has 0 cans and green otherwise. There are $2^{11} = 2048$ possible perceptual strategies. To allow for perceptual strategies to coevolve with foraging strategies, each robot has 11 more genes in its genome, which code the color that the robot sees for each quantity of cans. In the first generation the assignment of colors to the 11 genes is random. (Hoffman et al., 2015, p. 1488)

Hoffman et al. (2015) justify using non-monotonical fitness scores – (0,1,3,6,9,10,9,6,3,1,0) – by saying: "Such a nonlinear payoff function is quite common: Not enough water and one dies of thirst; too much and one drowns; somewhere in between is just right. Similarly for salt and a variety of other resources. Indeed, for organisms that must maintain homeostasis of a wide variety of variables, one can expect many nonmonotonic payoff functions" (Hoffman et al., 2015, p. 1486). It also needs to be noted that this Robby can only do six moves and not seven (as did Mitchell's): it can move north, east, west, south or random, or it can pick up a can, but it cannot stay put on the same space as Mitchell's Robby could do as well.

What happens in this algorithm is very similar to Mitchell's. Robbies start out with extremely wasteful strategies and end up mastering their environment. What is interesting is what happens with the perceptual strategy and the DNA of Robbies. They end up having one of the two variants of perceptual DNA. "In the first, squares are seen as *red* [italics Hoffman et al.] if they contain 0, 1, 9, or 10 cans, and as *green* [italics Hoffman et al.] otherwise. In the second, it is the reverse, with squares seen as *green* [italics Hoffman et al.] if they contain 0, 1, 9, or 10 cans, and as *red* [italics Hoffman et al.] otherwise" (Ibid.). This means that the organism sees the colour according to the amount of cans it needs for survival, which is information about the organism's fitness, and not colours according to the quantity of cans, which is information about the real world. Figure 9 shows both variants and Figure 10 shows what perceptual DNA of a Robby that sees the world isomorphically (perceptions attuned to "reality") looks like and what perceptual DNA of a Robby that sees the world non-isomorphically (perceptions attuned to fitness) looks like.

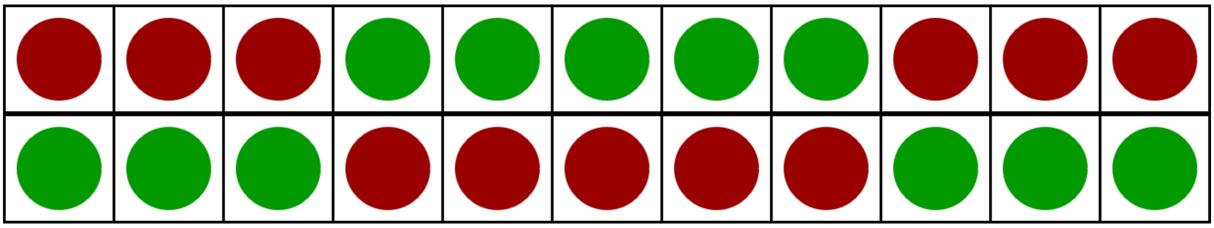

Figure 9. Two variants of non-isomorphic perceptual DNA.

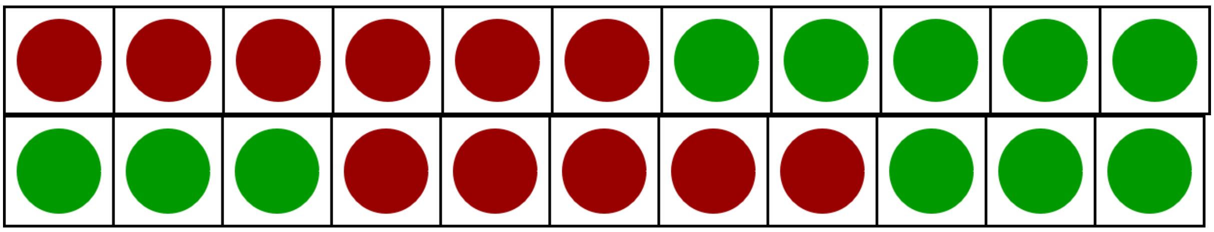

Figure 10. Top: isomorphic perceptual DNA. Bottom: non-isomorphic perceptual DNA.

To summarise: Hoffman's Robby associates the number of cans with two colours – red and green. It then has to learn to act according to those colours, assigned to the number of cans, in a way that brings it the most points. Through many generations, Robby devises a perceptual strategy that associates squares with high scoring profit as green (behind which is a certain quantity of soda cans) and squares with low scoring profit as red, or vice versa. Robby's perceptions are not isomorphic to the world out there, but to its internal dynamics.

This is the information that Hoffman et al. (2015) make available about their model. Apart from the major paper from 2015, Hoffman as the principal author of the theory has produced a number of papers where the GAs' role in ITP are referenced and conceptually described (Hoffman, 2009, 2011, 2018; Hoffman et al., 2015; Hoffman & Singh, 2012; Hoffman, Singh, & Mark, 2013). However, it seems that there is no complete report of computational, algorithmic and especially implementational data, therefore not being open source and having code on display for inspection. This prevents scrutiny, replication and further research. The various papers by Hoffman and colleagues contain scattered information. There are some descriptions of how the DNA is encoded, specifically that it encodes perceptual states and associates the number of cans to the two colours. The encoding of perceptual states is described and numbered. Some limited information is given on crossover, specifically that a pairing between two organisms produces two offsprings with randomly spliced original genes, where each offspring has a version of it. Regarding mutation, only that "a small amount of mutation is applied" (Hoffman et al., 2015, p. 1488) is given. The population number is set to 200, but no reason is listed. The stochastic parameters – the probability of mutation, the probability of individual organisms being selected for gene propagation – in general seem to be arbitrarily set, which is to a certain degree true for the ways genes crossover and mutate, and to the definitions of fitness evaluation and DNA encoding as well. This is important as by fine-tuning, changing the methods of recombination and mutation, and adapting fitness and encoding, different outcomes are produced. What is more, a complete reproduction in terms of algorithms and appropriate functions seems impossible. For example, what I struggled with was to understand how Mitchell's states of 0, 1 and 2 (no can, can and wall, respectively) are related to SEM and its colours, as these three possible states were kept (as was the whole foraging DNA). Since Hoffman et al. did not do anything about the wall, it was reasonable to assume that 2 was kept as a symbol or arbitrary number for the wall. Therefore, it seemed that 0 and 1 represent colour (either red or green, which is decided arbitrarily). In the next section, the process of reproduction will be described, noting what is missing from the description of SEM available online.

5.1.2 Reproduction of the Sensory-exclusive Model (SEM)

I managed to reproduce Hoffman et al.'s model in its general results – that what evolves is perceptual DNA and therefore perception that is attuned to fitness, not "truth" (see Figures 9 and 10). To be as transparent as possible to avoid problems that the unavailability of the model from the authors caused, the code for the reproduced model was uploaded to GitHub. It is available on the following link: https://github.com/TineKolenik/hoffman_reproduction. The link also appears in Appendix C. The reproduction of SEM and the model with enactivist presuppositions that will be presented in the next chapter are based on Ray Dino's "Robby – genetic algorithm" (Dino, 2015), licensed under the MIT License, where they reproduced Mitchell's GA model of Robby the Robot. For example, what needed to be added to Mitchell's Robby was a translating function that mapped squares with 0–10 cans to a colour 0 or 1. Therefore, if a Robby was in a situation with 5 cans in the northern space, 0 cans in its current space, 2 cans in its southern space, 3 cans in its western space and 9 cans in its eastern space, this had to be translated into a 5-digit number with values from 0 to 2. This is where the significance of the perceptual DNA lies, as it works as a dictionary for this translation. If the perceptual DNA was “11000000011” (where 0 and 1 are either green or red, and the index of the value represents the number of cans, so (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) cans, respectively), 5 cans were translated, by looking up the value under index 5 in the DNA, as 0, 0 cans as 1, 2 cans as 0 and so on. The translated 5-digit number was then looked up in the foraging strategy DNA to see which move is associated with it – this is how Robby worked and decided to make moves.

The reproduction of SEM always produces Robbies whose perceptual DNA reflects their fitness and not “truth” – they are therefore always non-isomorphic. However, depending on how I tweak the settings that quantify properties such as the population number, number of generations, mutation probability, etc., I get different evolutionary paths and (local) maxima in terms of scores for Robbies. The perceptual DNA is neither always the way Hoffman describes it, i.e. “11000000011” (where 0 and 1 are either green or red, and the index of the value represents the number of cans, so (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) cans, respectively). Sometimes, “10000000001” or “11100000111”[5] are produced. However, this still represents fitness-based, rather than “truth”-based perceptual DNA. The settings in Figure 11 seem to reliably produce “11000000011” or “00111111100”, the DNA that Hoffman et al. (2015) report.

```
POPULATION = 200

GENERATIONS = 500

LIFESPAN = 50

TRIES = 100

MUTATION = 300
```

Figure 11. Some of the arbitrarily set quantities and probabilities.

The settings represent the following:

a) POPULATION: sets the number of Robbies in a population.
b) GENERATIONS: sets the number of times Robbies go through the evolutionary process of fitness evaluation, crossover and mutation.
c) LIFESPAN: sets the number of steps Robbies take in their worlds.
d) TRIES: sets the number of worlds one Robby goes through in one generation before calculating average fitness, which determines the probability for crossover.
e) MUTATION: sets the denominator for the probability of a gene to be mutated, as in 1 in “mutation value” chance to be mutated.

These settings reliably produce “11000000011” and the graphs in Figure 12. They show the following: The top graph features the number of generations on the x-axis and the fitness score on the y-axis. The red colour shows the Robby with the minimum fitness score in a generation; the black colour shows the average fitness score of a generation; the green colour shows the Robby with the maximum fitness score in a generation. The bottom graph features the number

[5] This is only one possible representation of the DNA in such a form, it could be reversed in terms of 0s and 1s, e.g., 00011111000 – this is arbitrary.

of generations on the x-axis, and the fitness score variance, which will be beneficial in future analysis, is featured on the y-axis.

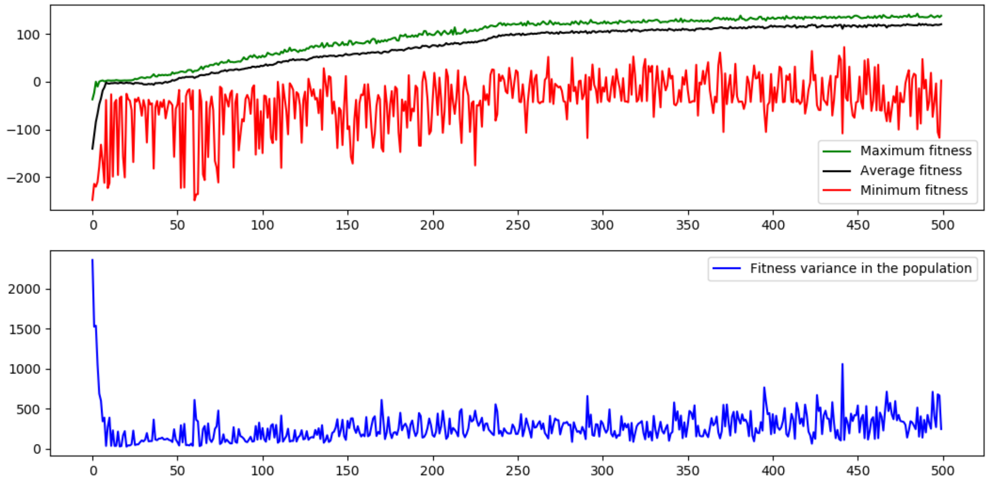

Figure 12. Two graphs denoting Robbies' progression through the evolutionary process. The top graph features the number of generations on the x-axis and the fitness score on the y-axis. The bottom graph features the number of generations on the x-axis and the fitness score variance on the y-axis.

Hoffman et al. (2015) did not include any such graphs to see Robbies' progression. This is the graph from Mark (2013), who initially came up with SEM:

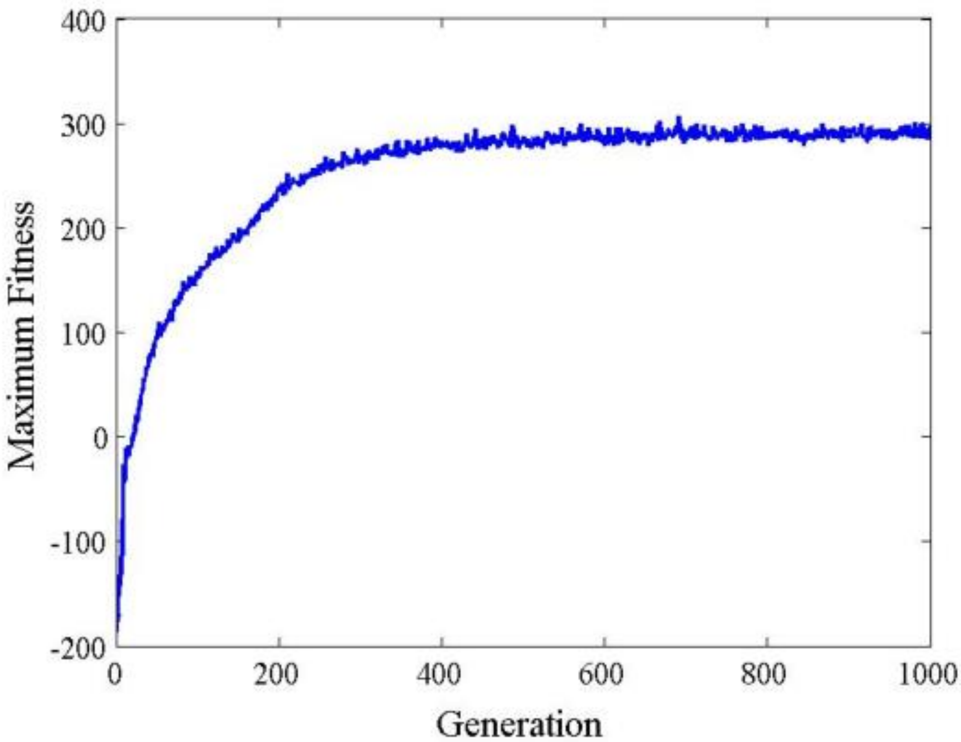

Figure 13. Graph from Mark (2013, p. 81) denoting Robbies' maximum fitness over generations.

Mark (2013) has no more significant information about SEM than Hoffman et al. (2015) had. Comparing Mark's graphs with mine, the difference in steepness and curvature of the slope as well as maximum fitness score is due to several factors, some known and some unknown. The first one is the number of steps that Robby takes (LIFESPAN setting). Mark's Robbies take 200 steps, while my Robbies take 50 steps. Other differences are due to other arbitrary settings, but there is a possibility that there are other, optimisation-related differences in the code. What is important is the perceptual DNA result, which is generally the same in the sense that it predictably ends up being non-isomorphic, i.e. fitness-based and not "truth"-based, while there might be deviation in which number of cans is labelled with which colour, as described before. For illustration, this is one example of the foraging strategy DNA: "00324004053153235050100045100220420250450150502200415405211501450550154315
52413345202345435355335015105355545112225415135455045145350355115114445105132051111254051551533453153124453053223324013552551153352251553340023420543
5253411140242115213".

Such a DNA can evolve alongside fitness-based perceptual DNA. The values correspond to the possible moves in this way: "0" means "move north", "1" means "move east", 2 means "move west", "3" mean "move south", "4" means "move random", "5" means "pick up". Indexes of these move values correspond to states by being translated from a base-10 to a base-3 number. For example, index 0, being a base-10 number, can be translated into a base-3 number, the number 00000. This corresponds to the state 00000, the 5-digit number that describes a scenario Robby is in (concretely where it is surrounded by something it perceives as five things that are the same colour). Base-3 of index 146, for example, is 12102. This represents another scenario for Robby. Therefore, according to the scenario it is in, it performs a move, represented by one of the values from 0 to 5.

Now that SEM has been arguably successfully reproduced and understood in-depth, the model's presuppositions, relevant to the endeavour of the thesis, can be explored.

5.2 (Epistemological) Presuppositions of the Sensory-exclusive Model (SEM)

The presuppositions that SEM possesses will be discerned from Hoffman et al.'s paper (2015) as well as from the model itself. In the paper, there are two revealing claims. The first one unveils Hoffman's and his colleagues' thoughts on the relationship between information processing and cognition: "The interface theory does not [deny information-processing]. Evidence for information processing is now overwhelming" (Hoffman et al., 2015, p. 1501). The second one reveals Hoffman et al.'s position on the relationship between perception and action: "Although the role of action is emphasized in sensorimotor and enactive approaches to perception […] (Noë 2006; Chemero 2009), our position differs in a crucial respect. In our view, having a perceptual experience does not require motor movements" (Hoffman et al., 2015, p. 1497). These two claims and the model's structure are central to the following analysis to discern (epistemological) presuppositions.

The core presupposition of Hoffman et al.'s model may be its most apparent one. It is manifested through the fact that the model is a computational model. It is based on computational principles, and Hoffman's team believes and treats the model's results as if they firmly tell us how the mind and cognition work as such. The core presupposition that can be discerned is that the mind and cognition work as a computer and are information-processing entities. Their first claim that the evidence for information-processing is overwhelming supports my conclusions. This firmly places the model in the cognitivist tradition. Hoffman et al. do not say that cognition "entails the manipulation of explicit symbolic representations of the state and behaviour of an objective external world" (Vernon, 2005, p. 6), but by taking the model's results seriously one has to come close to that, as Hoffman et al. do affirm that cognition (and perception) work as an information-processing machine. It may be that Hoffman et al. do not completely subscribe to the computational cognitivism (Putnam, 1968), but the second claim by Hoffman et al. supports my conclusions on their paradigm beliefs and the model's presuppositions.

Their second assertion, the claim that "having a perceptual experience does not require motor movements" (Hoffman et al., 2015, p. 1497), is especially revealing and gives further insight into the cognitivist bent of SEM. The authors themselves posit their view on the connection between perception and action in partial opposition to enactive and sensorimotor approaches. This is another presupposition that SEM carries. There are no significant sensorimotor loops, since the cognitivist idea the authors ascribe to does not deem it constitutive for cognition. Their statement hints that SEM does not lack a sensorimotor loop only due to its simplicity, but due to the ideas about cognition of the researchers. Hoffman et al., however, do not offer a deeper explanation of why they believe that visual experience does not require motor movements. This is also not apparent from the model, apart from the mere fact that it works in a predictable way and that it can be taken as evidence for ITP. SEM does contain an implicit connection between perception and action – they do interact and influence each other, as Robby's movement does result in its perception being changed from perceiving one set of spaces to another set of spaces; perception and action are also consequential and realised in two strictly separated, discretised steps. However, there is no necessary action to perceive at all in the first place and no feedback as such. The importance and the role of the sensorimotor loop, and the really deep connection between perception and action from the enactive perspective will be described in the next chapter where the sensorimotor model will be described.

Robby as the agent in the model has information of the true structures of the world and how to translate them into perceptions of two colours, even though it does not directly act on that information. This information is embedded into Robby's perceptual DNA. In this way, Robby is still an *a priori* realist to a degree that this cannot change, even though on another level it does evolve non-isomorphic perception. By playing the devil's advocate, it can be argued that perceptual DNA represents the eye structure that transforms substance from the objective, outside world (light waves) into something subjective to the agent itself (nerve impulses, vision and neural structures, etc.). However, the agent still seems to convert everything that it gets to its perceptions – there is no price to pay for perceiving, and Hoffman's Robby perceives everything around it. Robby does not have to decide what to perceive, to perceive only what pays off to perceive, to perceive only certain elements of the world and ignore elements of the world not important to the agent by not perceiving them. It is not clear whether this is due to the cognitive paradigm the authors ascribe to or whether it is due to the simplicity of the model. In any case, ambiguity of perceiving the world in this way is certainly possible to be embedded in Robby without significantly raising complexity, if at all.

The presuppositions in SEM are discussed in light of my own model, dubbed as the sensorimotor model (SMM), as all of the noted presuppositions are, to different extents, rectified and changed.

6 The Sensorimotor Model (SMM)

In this chapter, I am going to investigate whether different views on cognition, when modelled, affect the result in researching epistemological questions, in this case isomorphism between modelled perception and modelled reality. This will give some hints into what is important in computer modelling and what influences the final results. The insights will be valuable to evaluate GAs and computer modelling in general as methodology in cognitive science as natural epistemology. This will be done by presenting a model that is in many regards the same as SEM, yet with some differences in crucial paradigm presuppositions embedded in it.

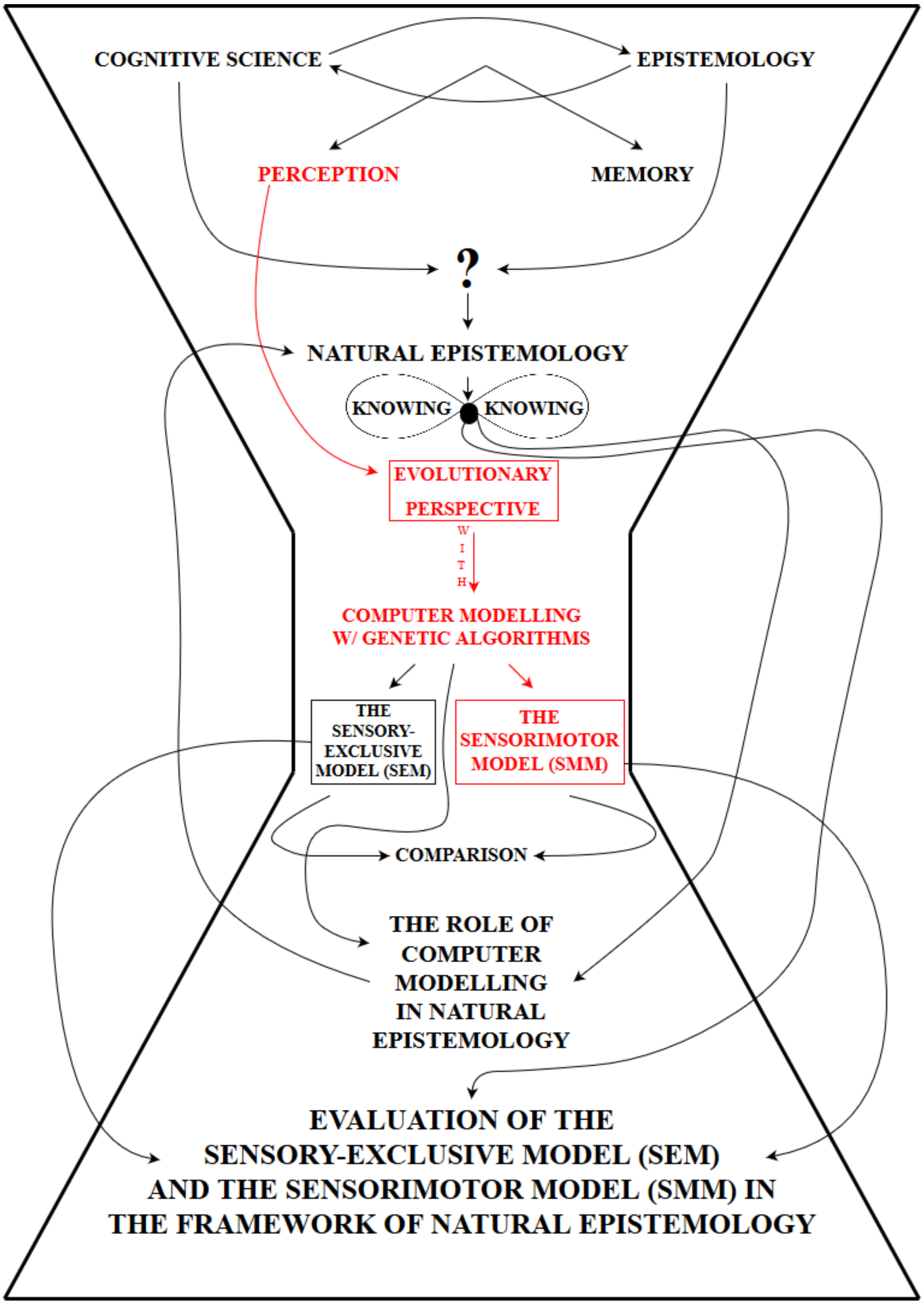

Figure 14. The position in the narrative schematic of the thesis, marked in red.

First, I will delve into enactivism and try to ascertain what can be modelled in general and what is appropriate to be modelled in the case of the sensorimotor model (SMM). I will look into computer models with enactive elements to look for precedence and see the different approaches to modelling enactivism.

6.1 (Epistemological) Presuppositions of the Sensorimotor Model (SMM): Enactivism

Enactivism, many times attributed to Varela, Thompson and Rosch's work *The Embodied Mind* from 1991 (Ward, Silverman & Villalobos, 2017), is an accumulation of a set of ideas that were at the time not new per se. Enactivism's foundations have been set throughout the 20^{th} century by such scholars as Piaget with his cognitive development theory (Piaget, Gruber, & Vonèche, 1995), Uexküll (1957) and his concept of Umwelt, Jonas (1966) with his biophilosophy, von Foerster (1962) with the second-order cybernetics movement and others. It can be seen as an intersection between biology (starting with Maturana's view on the question of "What is life?" and subsequent works, especially with Varela (Maturana & Varela, 1988)), cybernetics (shifting attention from material to processes, and later on shifting from observing systems to observing observers that observe systems) and existential phenomenology (from the likes of Edmund Husserl, Martin Heidegger and Maurice Merleau-Ponty). Some contemporary cognitive science theories also lean towards enactivist ideas, including Van Gelder's dynamicism (1995), predictive processing (Clark, 2013) and enactive approaches to visual perception (Gibson, 1986; O'Regan & Noë, 2001). However, enactivism is still a young paradigm, facing a lot of issues (lack of methods, non-unified view among scholars, trivialisation, etc.), making it not yet fully developed. This is especially true in comparison with cognitivism. Ward et al. (2017) roughly discern three main enactivist branches:

1. Autopoietic enactivism, which "emphasises and develops this attempt to ground cognition in the biodynamics of living systems" (Ibid., p. 369) with "a commitment to the strong continuity of life and mind—the view that the organisational structures and principles distinctive of mind are simply enriched versions of the structures and principles grounding life itself" (Ibid., p. 370), holds that "simple sensorimotor interactions between organism and environment bring with them a form of teleological directedness to the environment (as when the bacterium's activity aims at the sucrose-rich portions of its environment) and significance or value in the organism's environment (as when the chemical composition of the bacterium's environment becomes good for the bacterium in virtue of the dynamics of its embodiment)" (Ibid., p. 370).
2. Sensorimotor enactivism, which proposes "to account for the content and character of perception by appealing to sensorimotor contingencies: patterns of dependence obtaining between perception and exploratory activity" (Ibid., p. 371), claims that "a mobile perceiver has some control over the way in which sensory stimulation unfolds" (Ibid., p. 371).
3. Radical enactivism or Radically Enactive Cognition (Hutto & Myin, 2013), which aims to "unify anti-representationalist approaches to cognition" (Ibid., p. 372), claims that "intelligent behaviour can emerge in the absence of internal representations and mediating knowledge or understanding of the functioning of the system" (Ibid., p. 372).

Generally, enactivism can be seen as a non-reductive and non-functionalist naturalism (Di Paolo, Rohde, & De Jaegher, 2010; Thompson, 2010). Cognition is seen as embodied and embedded, as a continuously emerging product of a structural coupling between an active agent and the environment through the sensorimotor loop. This brings forth the agent's experiential world or Umwelt, the only accessible reality for the agent. The agent's cognition does therefore not exist to solve tasks, but rather to generate meaning for the agent in the form of enaction. Di Paolo et al. (2010) identify five core principles of enactivism: autonomy, sense-making, emergence, embodiment, and experience. These seem to underlie in all three branches of enactivism. The following section describes the five core principles of enactivism as proposed by Paolo et al. (2010).

6.1.1 The Five Core Principles of Enactivism

In this section, five key principles of enactivism are described: autonomy, sense-making, emergence, embodiment and experience. The description is based on Di Paolo et al. (2010) and Thompson (2010). The descriptions are meant to serve as building blocks that will be evaluated in terms of computability and suitability for my model.

6.1.1.1 Autonomy

Autonomy is, according to enactivism, an exclusive domain of living beings. It marks the ability of living organisms to set up their own laws to follow. This is a direct consequence (it could be argued they are equivalent) of organisms defining their own identities as distinct entities in the environment of precarious conditions. A system's identity has to be constantly self-generated; if it is specified by a designer, it follows the laws of the designer as well, and is consequently not autonomous (making autonomy of artificial systems, even though generally claimed otherwise, non-existent or trivial (Kolenik, 2016a)). An identity is generated "whenever a precarious network of dynamical processes becomes operationally closed" (Di Paolo et al., 2010, p. 38). The system is operationally closed when each of its processes is enabled or constituted by one of its processes and enables or constitutes another of its processes. Operational closure makes systems autonomous in regards to their environments as well (despite being coupled to them). This dissolves the prevalent view of agents being passive and only appropriately reacting to external stimuli, and makes them active in creating the laws of interactions with the environment, resulting in bringing forth or enacting their own world.

6.1.1.2 Sense-making

Autonomy, which results in an enaction of an Umwelt, is intricately connected with sense-making (or arguably the same thing). Sense-making describes the ability of organisms to create meaning in the world, to characterize what elements in the world are significant to it. This creation of meaning serves to maintain the identity of the organism that enabled this creation of meaning in the first place (which happens through interactions with the world, therefore through structural coupling and sensorimotor loops). Again, sense-making is active, not passive. As Di Paolo et al. say: "Natural cognitive systems are simply not in the business of accessing their world in order to build accurate pictures of it" (Di Paolo et al., 2010, p. 39). It is completely the opposite – by creation of meaning, by identifying significance in the world and transforming it in order to their needs, they enact their world. Di Paolo et al. (2010) imply that cognition is dynamical, biologically grounded sense-making.

6.1.1.3 Emergence

Emergence poses one of the most elusive concepts not only in enactivism, but in science in general. To break from the vagueness of using emergence whenever some magical step or leap from A to B cannot be explained, emergence in enactivism is grounded in a more pragmatic and biological phenomenon of self-organization. Both autonomy and sense-making are described as emergent phenomena. Autonomy is not a mere sum of all if its parts, and sense-making results in an Umwelt, which is again not a mere collection of certain processes. The biological groundedness of emergence is evident through a thoroughly enactivist concept of autopoiesis, which spawns life itself. Biological life can therefore be seen as the best example for emergence. Autopoiesis as the property that encapsulates self-generation, self-production and regeneration of a bounded system of molecules (e.g., a cell, the lowest level that exhibits autopoiesis) cannot be explained by studying various single elements of a cell (molecular components, energy and chemicals, proteins, DNA and RNA, the membrane, etc.). This is only the first step in trying to define what life is, as emergence works on many levels in organisms. Accepting emergence necessarily leads to questioning the reductionist approaches of studying various (cognitive) phenomena in single components and continuously lower levels of description (e.g., neuroscientific implications of cognition being bounded to the brain).

6.1.1.4 Embodiment

Embodiment is the most pragmatically successful concept in enactivism that made it into mainstream science, even though it has been in the process trivialized, mostly taking embodiment to mean that a cognitive agent's body influences their cognition (Kolenik, 2016b). It is therefore important to characterise it in the intended way. The body is a necessary component in a living being and defines the agent's sense-making, and therefore its Umwelt, to a degree that it defines its identity (taking a body from a living being would make it change its identity). The brain is not a control centre for the body, making the latter inferior, but an equal partner in sense-making.

6.1.1.5 Experience

Enactivism puts a lot of importance onto experience, a previously overlooked phenomenon, be it because it was thought as an epiphenomenon or because it cannot be described through reductionist scientific methods. Experience is an important window to the Umwelt itself and the apprehension of sense-making. Transformation of experience can be linked to the workings of sense-making. Di Paolo et al. also point to the correlating bodily changes that cause experiential changes and "isomorphisms" (Di Paolo et al., 2010, p. 44) between processes and experience. However, Di Paolo et al.'s descriptions of how experience fits into enactivism and its significance are particularly vague, which may be one of the reasons that the three branches of enactivism do not rely on experience as much as on other principles.

6.1.2 Possibilities for Modelling Enactivism

Determining what parts of enactivism can be modelled is a perplexing venture, as there seem to be many opposing views on the matter (Froese & Ziemke, 2009), which is to be expected, as there is no unified view on enactivism in the first place. My approach to exploring possibilities of modelling enactivism will be to examine the three branches of enactivism and the five core principles, analysing what has the predisposition to be computationally described. For help and justification, I will try to establish precedence by looking into examples of enactivist computer modelling. There are two questions concerning this examination, relating to my work: Can this part of enactivism be modelled? Is it sensible to include it in SMM?

6.1.2.1 Modelling Enactivisms and Their Principles

First, the key elements of the three branches of enactivism have to be discerned. These key elements have to be analysed in terms of computability as well as their suitability for SMM.

Autopoietic enactivism (AE), as Ward et al. (2017, p. 369) write, is based on the "closely related notion of autonomy," which "is emphasised—the way in which the self-sustaining biodynamics of autopoietic systems create both a distinction between an organism and its environment, and a domain of interactions that bear on the organism's prospects for survival." Furthermore, as described before, AE puts much focus on sensorimotor interactions between the organism and environment for successful cognitive behaviour and survival. I discern two key elements that are autopoietic enactivism's own: autopoiesis and autonomy. Sensorimotor interaction is vital as well, but not exclusive to AE; it is favoured by Thompson as well, who says that sensorimotor loops are necessary for sense-making (Thompson, 2007). It is important to determine whether autopoiesis and autonomy can be modelled.

Autopoiesis, meaning self-production and self-maintenance, is another topic where scientific community disagrees in regards to its potential for modelling. Varela and Maturana themselves were involved in a number of research projects, dubbed artificial chemistry, where they and their colleagues tried to model autopoietic organisation. This started in the 1970s (Maturana, Varela & Uribe, 1974), but later on Maturana became convinced that, by default, autopoiesis cannot be modelled, while Varela continued to work on it through the 1990s (Mingers, 1995). Varela claims that he and his colleagues have succeeded in modelling autopoiesis (McMullin & Varela, 1997) in very simple, low-level chemical processes with single particles. I am sceptical towards this claim and follow the thoughts of the likes of Maturana, Rosen (1991), Stewart (2000) and others, who argue that "the concept that the set of variables and the dynamic laws associated with a living organism result from the actual functioning of the organism itself" (Stewart, 2000, p. 157) cannot "be expressed mathematically" (Ibid.) This is because the set of variables and the dynamic laws always comes from the "the biomathematician [who] writes them down" and not "a living organism", "a system that is capable of (1) producing a semantic inscription and (2) interpreting that inscription" (Ibid.). The biomathematician is producing the semantic inscription and interpreting it for another system, while it is only capable of doing that for herself. I will therefore not try to model autopoiesis, and disbelieving that it can be modelled is not the only reason for that. Finding a way to include autopoiesis in regards to the level of SMM in terms of simplicity and abstractness would not be sensible. Even if I were convinced that autopoiesis can be modelled, it has so far only been modelled in very low-level, single particle situations.

Autonomy is, as mentioned before, closely related to autopoiesis, to the point where it is uncertain whether they can be effectively delineated. However, it has been, as a concept, more engaged with, especially in the artificial intelligence communities. Mostly it describes behavioural autonomy in relation to robots: "the robot is engineered so as to be able to interact with its environment without requiring ongoing human intervention" (Froese, Virgo & Izquierdo, 2007, p. 456), "any embodied system designed to satisfy internal or external goals by its own actions while in continuous long-term interaction with the environment in which it is situated" (Ibid.), and in rare cases to "autonomous systems are expected to survive in unknown and partially unpredictable environments by devising their own goals and finding out solutions to challenges that may arise" (Ibid., p. 457). Froese et al. (2007) argue that the concept of autonomy has been trivialised and that to have fully enactive systems, autonomy has to be constitutional. To be autonomous in the full enactive sense, a cognitive agent has to be capable of internally producing its own meaning, which consequently emerges goals. Not only that the meaning and goals must not be pre-programmed into the agent, the rules of the production of them must not be forced onto the agent. The agent itself should, through its internal operations, make its own rules of behaviour. This is highly problematic, as programming such an agent seems to be doomed from the start. The designer cannot escape the fact of pre-programming goals or the manner of producing the goals. Even if the goals were unpredictable, the designer would have to specify the rules under which these goals emerge. Without specifying the system at all, no code can be written in the first place. This is why autonomous agents, even though they seem abound in artificial intelligence (e.g., autonomous cars), cannot be said to be autonomous in the enactive sense. What is more, I believe that this constitutive autonomy cannot be modelled, and beyond very few attempts at vague conceptualisation of an artificial environment where autonomy could possibly emerge (Froese et al., 2007) and newer small-scale research endeavours, such as research in intrinsic motivation in artificial intelligence (Baldassarre et al., 2014), which cannot be reviewed yet in terms of their success, no models of autonomy seem to exist.

Sensorimotor enactivism (SE) places at its heart that perceptual experience is made possible by intricate operations of sensorimotor contingencies (Di Paolo et al., 2017). The basis for sensorimotor enactivism is laid out in *The Embodied Mind* by Varela, Thompson and Rosch (1991/2016), even though the book has been connected more to AE. Varela et al. say that the "enactive approach consists of two points: (1) perception consists in perceptually guided action and (2) cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided" (Varela et al. 1991, p. 173). These two points are at the core of SE, where the "focus of explanatory resources here is the primacy of action for cognition, the way in which motor variations induce (via the environment) sensory variations and the kind of regularities that emerge out of recurrent sensorimotor loops as constituting experience or organizing cognitive life" (Barandiaran, 2017, p. 413). One of the founders of SE, philosopher Alva Noë, writes that pushing sensorimotor enactivism to its extremes means that "perceiving is constituted by the exercise of a range of sensorimotor skills" (Noë, 2004, p. 90). The key element in SE is, as with AE, self-evident in the name – sensorimotor contingencies, their basic unit being the sensorimotor loop.

The sensorimotor loop is a circular process where "perception and action are inseparable, in the sense that they are co-dependent aspects" (Villalobos & Ward, 2015, p. 8). It "is manifested in the way in which motor variations induce (via the environment) sensory variations, and sensory changes induce (via internal processes) the agent to change the way it moves" (Di Paolo et al., 2017, p. 17). The sensorimotor loop seems to be one of the most accessible elements of enactivism for computer modelling (Di Paolo et al., 2017; Jug et al., 2018; Froese & Ziemke, 2009), probably as it is widely used outside of the strict framework of enactivism. It is therefore the main candidate for including it in my model, especially as it goes straight against Hoffman et al.'s claim that is evident in their model that "having a perceptual experience does not require motor movements" (Hoffman et al., 2015, p. 1497).

Radical enactivism or Radically Enactive Cognition (REC) holds as its "most central and important negative claim […] its denial that all forms of mental activity depend on the construction of internal models of worldly properties and states of affairs by means of representing its various features on the basis of retrieved information" (Hutto, 2013, p. 3), presenting itself as a fully-pledged and the most extreme "anti-representationalist paradigm" (Ibid., p. 10). It is difficult to find key elements of REC that are its own. This may be true because REC is not considered as a branch of enactivism by its founder Daniel D. Hutto, who says that "REC is not an alternative version of enactivism with distinct explanatory tools in its own right" (Hutto, 2017, p. 379), that "REC never stands alone" (Ibid.) and that "REC's aim is to radicalize existing versions of enactivism and related explanatory accounts" (Ibid.). Its dire anti-representationalism could be considered as a key element. However, it seems that anti-representationalism is not compatible with the question of whether perception is non-isomorphic or isomorphic. I do not know what, if not representations in the model, could be compared to the modelled world and deemed as isomorphic or not. I will therefore be agnostic to this extreme version of enactivism as proposed by Hutto. Trying to model non-representationalism in terms of my model does not seem sensible, and there is precedence that there are few, if any simulation models that try to model cognition in such a way. Brooks' real-life robots (1991) are presented as having no representations and running only on pure reflexive multiple-level sensorimotor loops. With many distributed levels of such loops, Brooks' robots exhibit intelligent behaviour, such as collecting cans in the lab and throwing them in the trash. Brooks patented the robotic system and went on to establish an extremely successful company – iRobot, and its intelligent vacuum cleaners are based on the same principle (e.g., if obstacle is hit, then change direction). However, these principles of non-representationalism have not been modelled in areas outside of rather simplistic perception and action loops.

In these three branches of enactivism, three core principles were evaluated in their feasibility for being computationally modelled. Autopoiesis and autonomy were deemed inappropriate for modelling, while sensorimotor contingencies seem to be modellable as well as a major part of enactivism, which would shift SEM away from the cognitivist framework. The two remaining core principles are emergence and experience. Being the most vaguely described of all the principles as well as not being focused on in the three branches of enactivism, I will not delve into them too much. Emergence seems to be applicable to almost anything – e.g., there is emergence in SEM in terms of its emergent intelligent behaviour that is evolved. Experience is especially indistinct in terms of description – Hoffman et al. (2015) use experience as the final stage of perception of their model's Robbies, and some argue that it may be potentially computationally modelled (Chrisley & Parthemore, 2007), while others argue that experience may only occur in living beings (Thompson, 2010).

In the next section, the transition from SEM to SMM will be described, where modification takes place so that SMM includes certain enactivist presuppositions, namely sensorimotor contingencies in connection to visual experience. The Python code for the model is available on the following link: https://github.com/TineKolenik/hoffman_enactivist_upgrade. The link also appears in Appendix C.

6.2 The Sensorimotor Model (SMM) Design

SMM will differ from SEM in one crucial area: the inclusion of the sensorimotor loop. The inclusion of the sensorimotor loop has somewhat wide implications in that it really does represent a paradigm shift in viewing cognition and its implementation, as there are seemingly significant changes to the code. Riegler et al. (2013, p. 4) say that "enrich[ing] [the cognitivist] concept with insights from constructivist sensorimotor contingency theory as a mode of world exploration through (local and communicated) action" is a valid approach to paradigmically (the paradigm being enactivism) relevant computer models.

SEM follows the general cognitivist separation of perception and generation of behaviour. SEM's Robby gets perceptual experience by the virtue of doing nothing – there is no necessary acting to perceive, as Robby can stand still, do no actions or moves and still know its state. The state and the scenario it is in is therefore explicated without the need for its active involvement in it. To implement the sensorimotor loop in the enactive sense, a loop cannot be simply added. Robby's perception has to be changed for sensorimotor loop to have a pivotal role in Robby's involvement with the world.

SEM's Robby has always, regardless of any of its possible actions, access to the information about the scenario it is in. This means that it gets information about the space it is in and the spaces north, south, east and west of it. The "true" information about the spaces is associated with perception – it perceives their colour. The sensorimotor loop in SMM's Robby was implemented by making Robby's field of perception more limited and completely dependent on its decision where to turn and look. This means that Robby does not perceive anything around it unless it actively turns and looks in a chosen direction, choosing between looking north, south, east or west – it then perceives the colour of the space it is looking at. By default, it always perceives the position it is in. To make the loop really influential in Robby's subsequent behaviour, Robby's decision where to look is the last step in the process. My Robby's loop of life is therefore the following:

1. Depending on where Robby is looking at, perceive the space's colour.
2. Make a move (move north, east, west, south or random, or pick up a can) depending on what Robby sees in the direction it is looking at and the space it is standing on.
3. Decide towards which space to turn to, which will be perceived in the reiteration of the process in step 1.

In step 1, Robby's perception of the space and its colour are completely dependent on Robby's decision where to turn to, which happened in the previous loop in step 3. Where to turn to is dependent on its perception of that space it is turned to. This is not only an abstract explanation of what happens, it is also how it is directly implemented in the code.

SEM's Robby's foraging strategy DNA consisted of the scenario of the five spaces and the associated move. For SMM's Robby's foraging strategy DNA, a modification took place. It is not the whole scenario that Robby acts upon, but only what it perceives from the direction it is turned to. This is associated not only with one of the six moves, but a turn as well. Instead of the association between the index in the DNA and the value 0–5 for the move, I made a tuple with (move, turn) pairs. The turn has a value of 0–3: "0" means "turn north, "1" means "turn south", "2" means "turn west" and "3" means "turn east". It may therefore be ill-suited to still talk about a process with steps 1–3, as 3 is not a sharp end of the process for Robby, as it naturally leads to step 1 again, making it a circular behaviour where it is impossible to delineate between the start and the end, between initial perception and action or vice versa. The concept of the enactive sensorimotor loop represents exactly that.

6.3 Results of the Sensorimotor Model (SMM)

By implementing a key enactivist element into SMM, the evolutionary path of its Robbies somewhat differed from SEM's Robbies. Not in the final result, as the Robbies in the final generation of SMM all possess fitness-based, non-isomorphic perceptual DNA, specifically "00111111100" for the graph in Figure 16, but in the overall progress. The final (local) maximum was approximately the same. The comparison between SEM and SMM is discussed a bit more in the next chapter, "Comparison and additional experiments".

Figure 15 shows the various settings for SMM, which were the same as in the case of the reproduced SEM. Figure 16 shows the graph of Robbies' progress in minimum, average and maximum fitness scores over 500 generations.

```
POPULATION = 200
GENERATIONS = 500
LIFESPAN = 50
TRIES = 100
MUTATION = 300
```

Figure 15. Some arbitrary settings for SMM. They are the same as the settings in the reproduced SEM for them to be comparable.

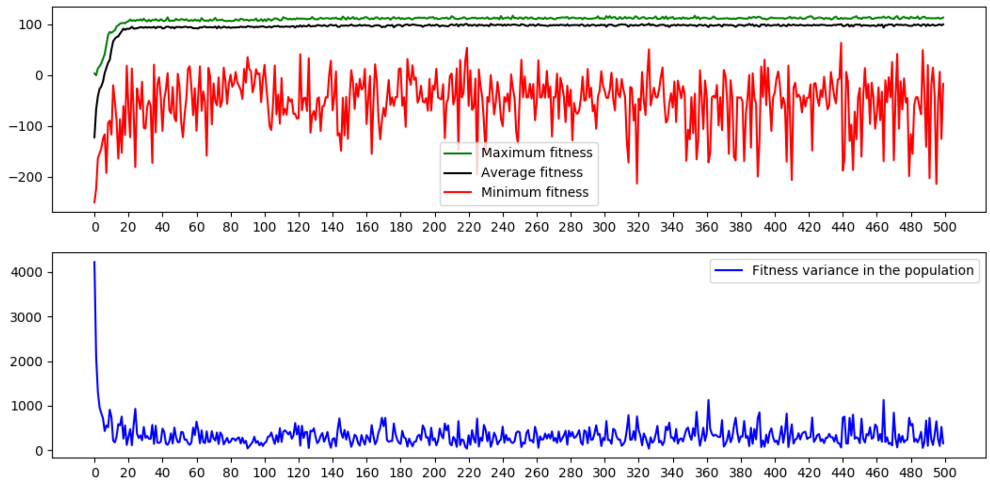

Figure 16. Two graphs denoting Robbies' progression through the evolutionary process. The top graph features the number of generations on the x-axis and the fitness score on the y-axis. The bottom graph features the number of generations on the x-axis and the fitness score variance on the y-axis.

7 Comparison of the Models and Additional Experiments

In this chapter, SEM and SMM will be compared more in depth. Then, additional experiments with the models will be presented.

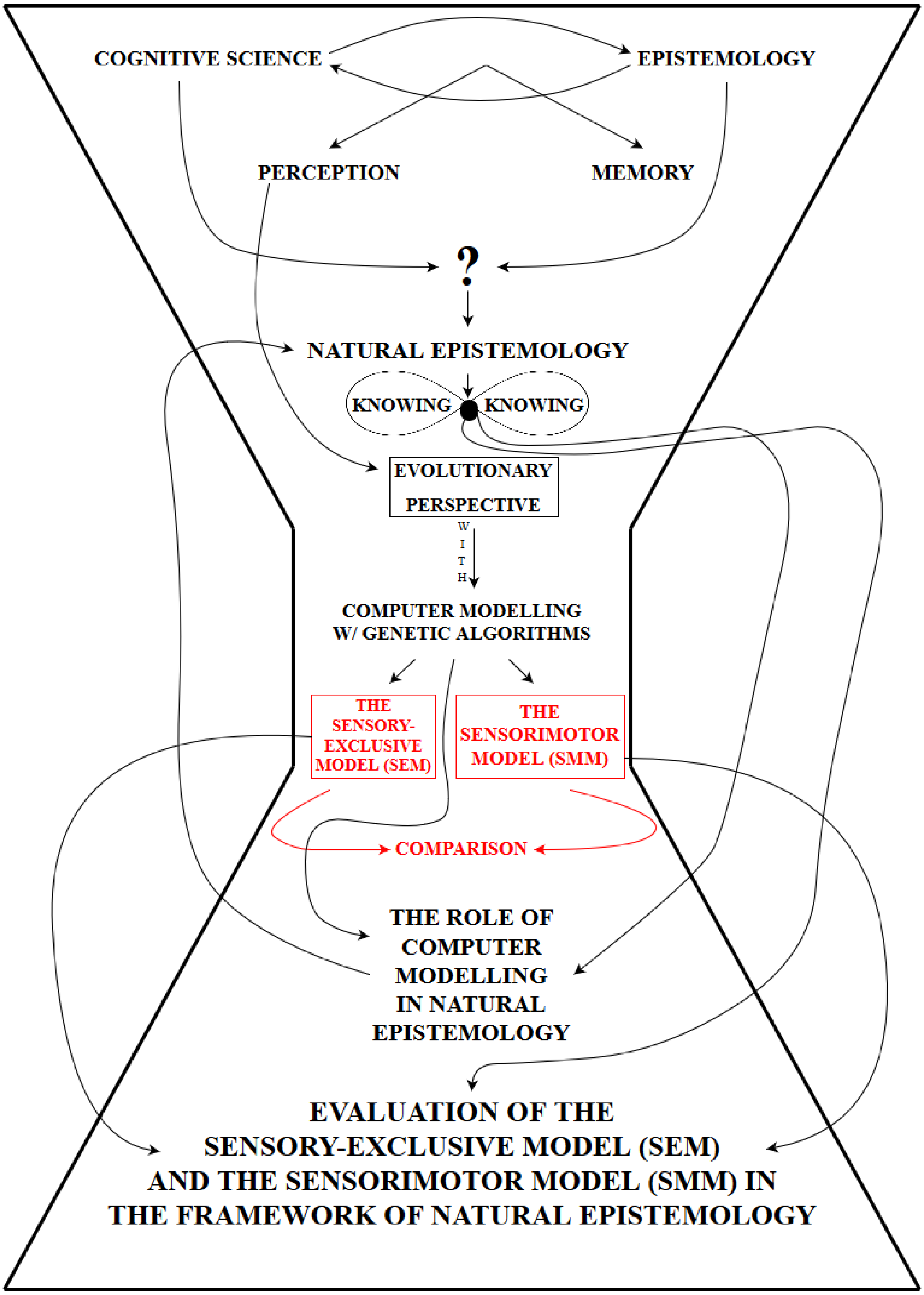

Figure 17. The position in the narrative schematic of the thesis, marked in red.

7.1 Comparison of the Sensory-exclusive Model (SEM) and the Sensorimotor Model (SMM)

Similarities between SEM and SMM start in the ideas upon which and for which they were constructed. Let's take a look at ITP and its analogy with the computer interface:

> [T]he interface theory of perception says that the relationship between our perceptions and reality is analogous to the relationship between a desktop interface and a computer.
>
> A desktop interface makes it easy to use the computer. To delete or copy files, for instance, one simply needs to drag icons around on the desktop.
>
> But a desktop interface does not make it easy to know the true structure of a computer—its transistors, circuits, voltages, magnetic fields, firmware, and software. Indeed, it's in part by hiding this complex structure that the desktop makes it easier to use the computer. Why? Because if you were forced to be aware of the true facts about circuits, voltages, and magnetic fields, when your goal was simply to edit a photo or write a paper, you would be wasting time, memory, and energy on truths of no relevance to accomplishing your goal.
>
> In similar fashion, says the interface theory of perception, our perceptions have been shaped by natural selection to make it easier for us to act effectively in the world, so that we can survive and reproduce (or, more accurately, so that our genes can survive and reproduce). Our perceptions have not been shaped to make it easy to know the true structure of the world but instead to hide its complexity.
>
> Our perception of space-time is analogous to the desktop, and our perception of objects and their properties is analogous to the icons on the desktop. Just as the language of desktops and icons is the wrong language for describing the true structure of the computer, so also the language of space-time and physical objects is the wrong language for describing the true structure of the objective world. A blue and rectangular icon on a desktop does not represent that something in the computer is blue and rectangular. Not because the icon is false or misleading or illusory, but because the icon is there to help you use the computer, not to distract you with irrelevant details about its innards. (Hoffman et al., 2015, p. 1488)

This telling analogy is markedly similar to the analogy Maturana and Varela, two founding fathers of enactivism, made at various times when describing cognitive agents' access and relationship to the outside world. Maturana (1978) makes an analogy using a pilot and a plane's cockpit:

> Let us consider what happens in instrumental flight. The pilot is isolated from the outside world; all he can do is manipulate the instruments of the plane according to a certain path of change in their readings. When the pilot comes out of the plane, however, his wife and friends embrace him with Joy and tell him: 'What a wonderful landing you made; we were afraid, because of the heavy fog.' But the pilot answers in surprise: 'Flight? Landing? What do you mean? I did not fly or land; I only manipulated certain internal relations of the plane in order to obtain a particular sequence of readings in a set of instruments.' All that took place in the plane was determined by the structure

> of the plane and the pilot, and was independent of the nature of the medium that produced the perturbations compensated for by the dynamics of states of the plane: flight and landing are irrelevant for the internal dynamics of the plane. However, from the point of view of the observer, the internal dynamics of the plane results in a flight only if in that respect the structure of the plane matches the structure of the medium; otherwise it does not, even if in the nonmatching medium the internal dynamics of states of the plane is indistinguishable from the internal dynamics of states the plane under observed flight. It follows that since the dynamics of states of an organism, or of a nervous system, or of any dynamic system, is always determined by the structure of the system, adequate behavior is necessarily only the result of a structural matching between organism (dynamic system) and medium. (Maturana, 1978, pp. 32–33)

The main parallel between the two analogies is the parallel between Hoffman's interface and Maturana's plane. Both seem to be windows to the outside objective world in terms of successful behaviours of the user (in Hoffman's case) and the pilot (in Maturana's case), guiding them through their structure in a way that is accessible to the user and to the pilot. Maturana and Varela use a very similar analogy again in their seminal book *The Tree of Knowledge* (1988), a spiritual predecessor to the *The Embodied Mind* (1991/2016), where they exchange a plane and a pilot for a submarine and a navigator, who, after being congratulated to have perfectly manoeuvred her submarine to the shore, exclaims: "All I did was push some levers and turn knobs and make certain relationships between indicators as I operated the levers and knobs. It was all done in a prescribed sequence which I'm used to. I didn't do any special maneuver, and on top of that, you talk to me about a submarine" (Maturana & Varela, 1988, pp. 136–137). In Maturana's and Varela's story, the navigator has no concept of anything else outside of the submarine's instruments. She does not know what the real, objective world is like – even more, she thinks the inside of the submarine is the world as such. This is parallel to Maturana's pilot's cockpit and Hoffman's computer user. In these examples, an outside observer is present, who is able to verify what the structure of the outside world is like. In epistemology, this is not possible, which will be addressed when the value of computer modelling will be discussed in chapters 8 and 9.

The analogies are reflected in SEM and SMM. In both of them, the outcomes of the evolved perceptual DNAs being curtailed to fitness ("00111111100") rather than "truth" ("00000011111") are analogous to the analogies where agents have a particular relationship with the outside, objective world – their perceptions serve evolutionary fitness, guide behaviour to be successful, but are not isomorphic to the "truth", to the objective world.

Comparing the models further, the fitness scores after 500 generations are extremely similar. Figure 18 shows graphs of the fitness scores over generations from both models' Robbies.

What can be seen is that Robbies from both models reach approximately the same maximum fitness score, around 120, but Robbies from SMM reach it much faster. This is an interesting phenomenon to investigate, as it seems as if the sensorimotor loop worked as an optimisation manoeuvre. This seems to be a novel discovery regarding sensorimotor loops and their application in artificial agents, used for researching cognitive phenomena. The next section describes how the sensorimotor loop works as an optimisation manoeuvre.

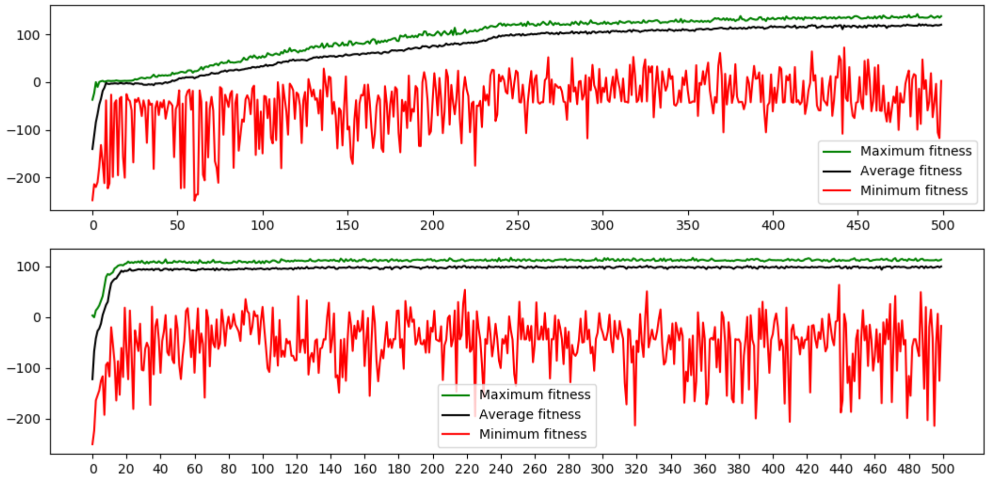

Figure 18. Top graph represents SEM's Robbies' fitness score over 500 generations. Bottom graph represents SMM's Robbies' fitness score over 500 generations.

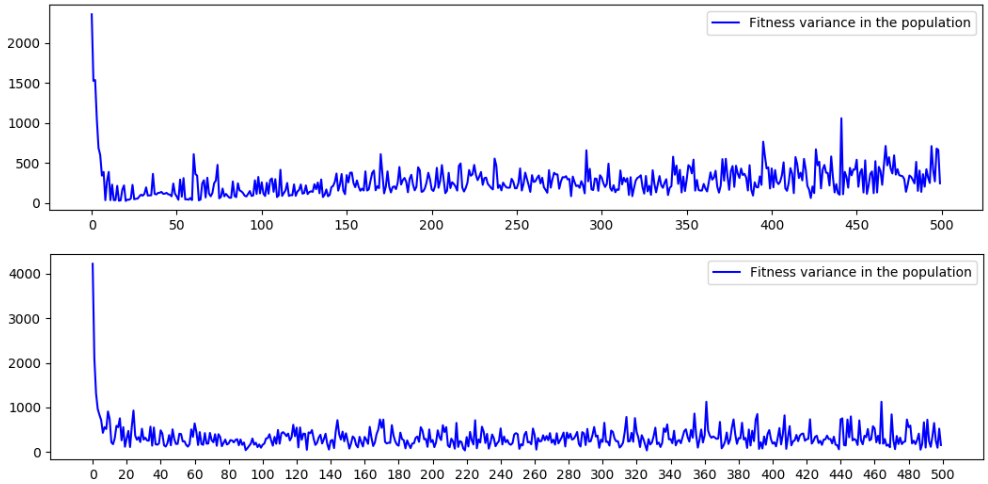

Figure 19. In the top graph, fitness score variance in a population is plotted over generations in Hoffman et al.'s model. In the bottom graph, fitness score variance in a population is plotted over generations in SMM.

Since I expanded the foraging strategy DNA by making pairs of (move, turn) out of single values with only (move), this means that the DNA complexity is higher. With higher complexity, there are more possibilities of different individuals in terms of their foraging strategy DNA when making the initial population. This means that there is a higher variance in terms of individuals at the beginning, meaning higher variance in terms of fitness scores. This can be seen in Figure 19. In the top graph, fitness score variance in a population is plotted over generations in SEM. In the bottom graph, fitness score variance in a population is plotted over generations in SMM. In SEM, starting variance in fitness score is below 2500. In SMM, it is above 4000. I hypothesise that this is due to the first generated population in SMM having more individuals with very high fitness scores as well as those with very low fitness scores. By looking at minimum and maximum fitness scores in the starting populations, which can be seen from Figure 18, it can be seen that in SMM, the best individual's fitness score is considerably higher than the best individual's fitness score in SEM, by approximately 35 points. This best individual in SMM bootstraps evolution by having a big probability for being selected into the mating pool for crossover. This results in a much quicker convergence towards the local maximum fitness score, which is, in SMM, reached after about 20 generations, while in SEM it is reached after 250 generations.

In the next section, additional experiments with both models will be presented, where organisms are implemented with fixed perceptual strategies. I experimented with two perceptual strategies – "truth"-based and fitness-based one. In the first experiment, the perceptual DNA was "00000111111" ("truth"-based, encoding information about the quantity of sodas in the outside world), while in the second it was "00001110000" (fitness-based, encoding information about the quality, about the survival needs of the organism). The perceptual DNA was fixed, meaning that all organisms in the first generation started with it and that it did neither crossover (even if it did, it would stay the same) nor did it mutate.

7.2 Additional Experiments with Fixed Perceptual Strategies

Four additional experiments were made with SEM and SMM to further examine legitimacy of fitness-based, non-isomorphic perception prevailing over "truth"-based, isomorphic perception. The models' Robbies were implemented with the matrix coding the mapping from the external world to colours that was constant, unchanged neither by crossover nor by mutation. Four experiments were run:

0. SEM was implemented with a fixed isomorphic perceptual strategy "00000111111" and run for 500 generations.
1. SEM model was implemented with a fixed non-isomorphic perceptual strategy "00001110000" and run for 500 generations.
2. SMM was implemented with a fixed isomorphic perceptual strategy "00000111111" and run for 500 generations.
3. SMM was implemented with a fixed non-isomorphic perceptual strategy "00001110000" and run for 500 generations.

Figure 20 shows the settings for arbitrary parameters, which were the same for all of the model runs and are the same as were for the previous analyses. Mutation is still present as it still occurs in the foraging DNA.

```
POPULATION = 200

GENERATIONS = 500

LIFESPAN = 50

TRIES = 100

MUTATION = 300
```

Figure 20. Arbitrarily set number for various settings for the four experiments.

Figure 21, 22, 23 and 24 show graphs for SEM with a fixed isomorphic perceptual strategy, SEM with a fixed non-isomorphic perceptual strategy, SMM with a fixed isomorphic perceptual strategy and SMM with a fixed non-isomorphic perceptual strategy, respectively.

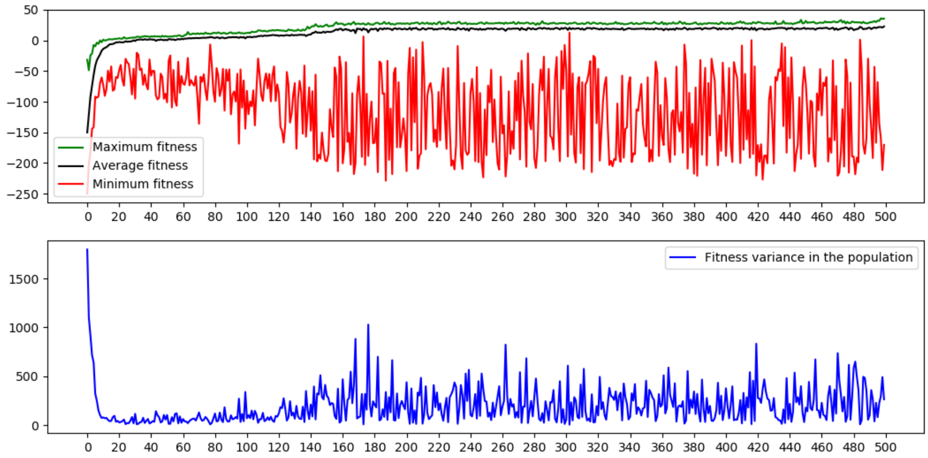

Figure 21. SEM with a fixed isomorphic (realist) perceptual strategy. The top graph shows the fitness score over generations, the bottom graph shows fitness score variance over generations.

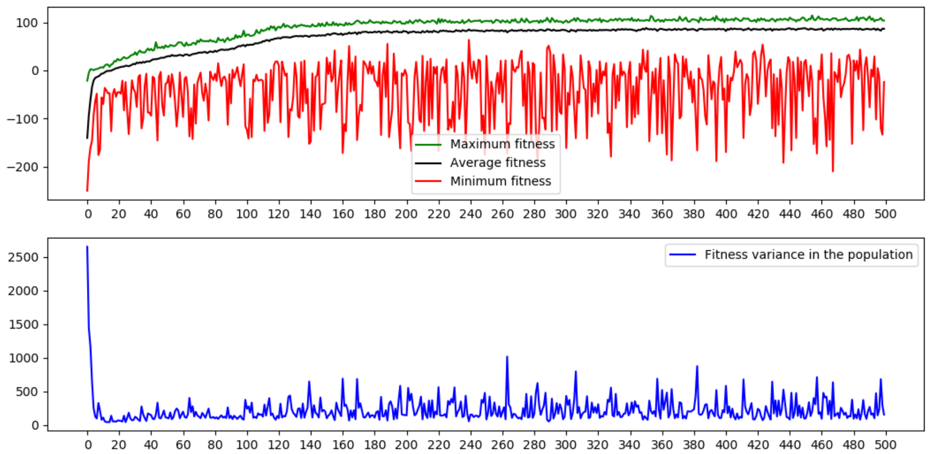

Figure 22. SEM with a fixed non-isomorphic (constructivist) perceptual strategy. The top graph shows the fitness score over generations, the bottom graph shows fitness score variance over generations.

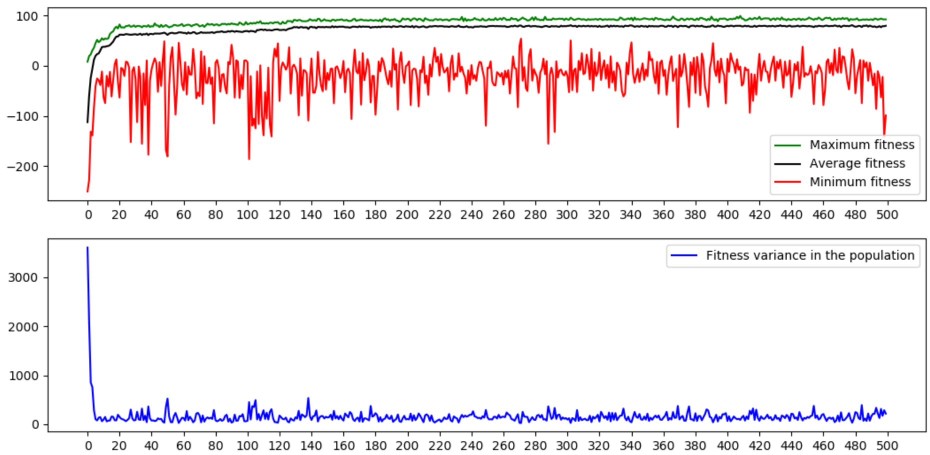

Figure 23. SMM with a fixed isomorphic (realist) perceptual strategy. The top graph shows the fitness score over generations, the bottom graph shows fitness score variance over generations.

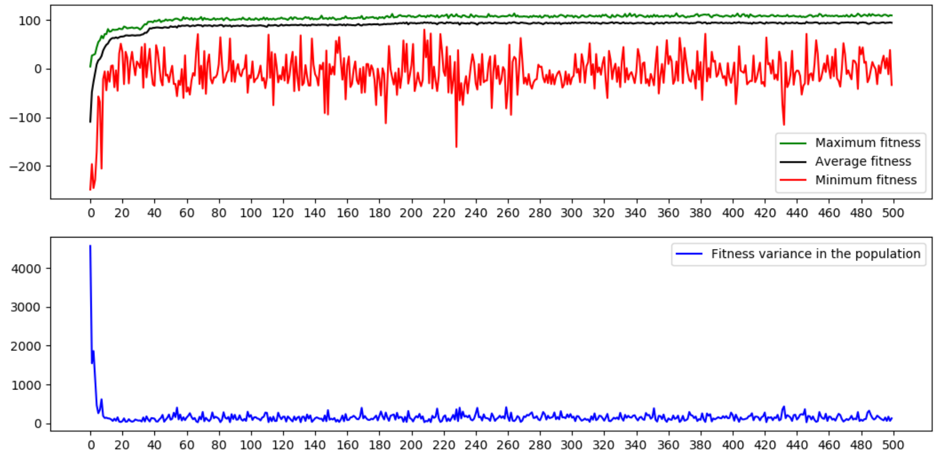

Figure 24. SMM with a fixed non-isomorphic (constructivist) perceptual strategy. The top graph shows the fitness score over generations, the bottom graph shows fitness score variance over generations.

Experiments mostly yielded nothing out of the usual. Both models with non-isomorphic perceptual strategies scored similarly and went above 100 fitness points, similar to the original models without fixed, but evolving perceptual strategies. The slope of SEM’s two graphs compared to SMM’s are again to be expected – the same happened in the models with evolving strategies. The same goes for variance. What it is unexpected is that Robbies with isomorphic perceptual strategies in SMM score around 90 fitness points, which is a lot higher than Robbies with the same perceptual strategy in SEM. This might be again due to the varying variance and higher scoring individuals in SMM, where the sensorimotor loop works as an optimiser.

Further experiments yielded results that were expected, and showed that the fitness-based, non-isomorphic perceptual strategy makes Robbies more successful in picking up soda cans and navigating the modelled world.

The work on computer modelling for this thesis has given me an insight into how the researcher can influence the results by making decisions regarding arbitrary pieces of the model, be it fine-tuning the settings or code implementation. What is more, the ideas about the mind, which wildly differ between different paradigms, are implemented into models at the researchers’ whims. The listed insights from hands-on computer modelling offer an informed perspective that is necessary for a detailed and distinct investigation into synthetic methodology as a method of natural epistemology. In the next two chapters (8 and 9), computer modelling in general as well as the two models specifically will be examined in terms of their potentials, limitations, implications and overall value for natural epistemology.

8 Examination of Computer Modelling as a Method of Natural Epistemology

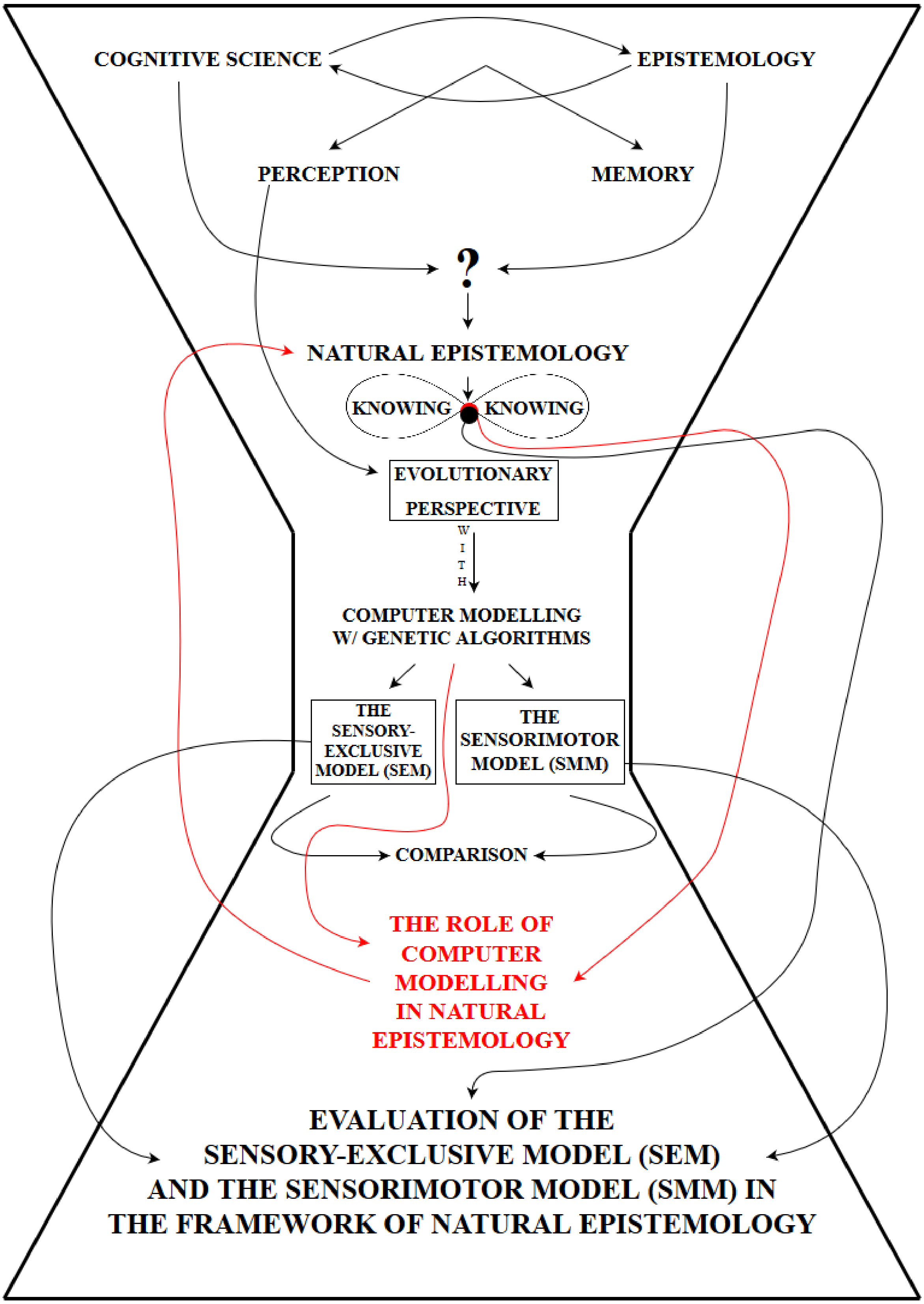

Figure 25. The position in the narrative schematic of the thesis, marked in red.

The history of cognitive science and its prevailing cognitivist leanings have made the term "computational" very loaded to the degree that it may have become ineffectual in theoretical contemplations. Riegler, Stewart and Ziemke (2013) agree that "the concept of 'computational model' is ambiguous and depends on whether it is used in the sense of the computational(ist) theory of mind or simply as a tool" (Riegler et al., 2013, p. 1). They argue that these two different senses in which "computational" can be termed have to be strictly delineated and carefully operated with:

> The first sense of the term is that the processes being modeled are conceived of as being themselves computational in nature. This is the sense that is deliberately, honestly and openly taken on board by the proponents of the classical 'computational theory of mind' (CtM) – notably by Jerry Fodor and Noam Chomsky, who were historically the prime advocates of this view. While this use of the term is perfectly legitimate, it does not mean that the CtM is above criticism; quite the reverse, in fact, because its very legitimacy lies principally in the fact that those who espouse it put themselves on the line, and lay themselves open to honest, rational criticism.
>
> The second sense of the term is that the author has used a computer as a tool to express a vision, to render certain aspects or properties of his model salient, to make them graspable by relatively direct intuition. This second sense of the term is quite different from the first. It is not (just) 'wider' or 'looser': that would invite an assimilation of the two, which would lead us back to the very conflation and confusion that we are trying to avoid. (Riegler et al., 2013, p. 2)

However, this is only one conundrum when trying to theorise precisely about computation. Even if ascribing to CtM, the question remains whether any model where the world is simulated can be thought of as computational in the first sense. This means that even if a part of the model, the agent, is faithfully modelled, a large part, the world around the agent, is not. Why is that? Many natural phenomena, other than those related to living organisms, can be, to different degrees, accurately described with mathematical formulas. These situations can be modelled – e.g., the behaviour of winds can be described with differential equations, simulated and the result will be a useful approximation (Zárate-Miñano, Anghel, & Milano, 2013). However, as opposed to the mind, no one believes that the winds are doing computations in and for their behaviours (Riegler et al., 2013). The idea is therefore that living organisms perform computations which are expressed by their behaviour in the world, while non-living nature simply does not do that. What happens then when the two are interacting – the modelled living organism and the modelled environment? If one is supposed to represent the true nature of the world (computational workings of living beings), while the other is supposed to represent approximate description of how the world works through mathematics, the ordeal seems to work against the first sense in which Riegler et al. (2013) describe 'computational' – "processes being modeled are conceived of as being themselves computational in nature" (Ibid., p. 2) – as the situation between a modelled agent and a modelled world would need a definition that is "'wider' or 'looser'" (Ibid., p. 2), which is problematic as "that would invite an assimilation of the two, which would lead us back to the very conflation and confusion that we are trying to avoid" (Ibid., p. 2). But there is more to this dichotomy between modelling the mind and modelling the world, and it is firmly connected to what 'computational' means.

Riegler (2001) identified one such dichotomy and labelled it as PacMan Syndrome:

> [I]t is useless to implement agents in artificial intelligence or artificial life with an a priori defined set of concepts, and to claim they were 'intelligent'. […] Artificial agents interact with anthropomorphically defined entities, such as ‘food’ and ‘enemy’, which make sense only to the programmer of the system. No attention is paid to questions like: How have organisms arrived at the idea that something is a source of food? How do they ‘know’ that another creature is a dangerous opponent? Predators do not have signs on their backs saying ‘I'm your enemy’. Even if this were the case — how would cognitive beings have learned to understand the meaning of those labels? (Riegler, 2001, pp. 4–5)

Riegler describes how agents are being modelled in the same way as the world is – by approximating observable behaviour from a third-person point of view rather than creating minds that would truly conform to the first sense of “computational”. The problem is therefore not in creating agents through paradigms that conform to CtM, but rather a much more puzzling dichotomy. This dichotomy seems to be stronger than Riegler et al.’s delineation between the two senses of ‘computational’. It is, as Füllsack (2013) implies, that living organisms self-create knowledge and therefore behaviour, while the non-living natural world does not. He is somewhat pessimistic in his brief thoughts on modelling self-creation of knowledge by artificial agents instead of their creators, saying that computer modelling can support the case for such ideas, but can never succeed in definitively producing it (Füllsack, 2013). This is somewhat similar to what Peschl (1991a) argues with his ideas on what he labels ‘computational neuroepistemology’. Peschl claims that “it is the designer who plays the role of evolution he/she designs the network architecture and the structure of the artificial cognitive system” (Peschl, 1991a, p. 2213), but that an interdisciplinary discourse that integrates “epistemological, connectionist as well as neuroscientific aspects” (Ibid.) to research knowledge development can be established by following certain conditions. He lists two: “1) the artificial neural network has to be physically embedded into its environment; this means that the communication between the system and its environment takes place via effectors and sensors - no symbols are involved in this process of interaction, 2) a recurrent topology which ensures a non-linear and non-trivial (‘intentional’) behavior” (Peschl, 1991a, p. 2210).

Undertaking ‘computational’ matters means that one is describing systems from a third-point of view, and it is this that seems to be the major issue. Even if one ascribes to the paradigm that believes the mind functions on performing symbolic computations with arbitrary symbols, the mind cannot be computationally described by a third-person observer by inputting knowledge, goals, the rules of the game of living, etc. into the agent. The researcher as the ultimate creator cannot faithfully create agents, even if she can approximately describe behaviours, be it of living or non-living nature. Kjellman (2013, p. i) figured out something similar in his doctoral dissertation: “The efforts of computer modelling and simulation analysis revealed a pronounced observer-dependency regarding investigation.” Observer-dependency will be described at length in the section on programming genetic algorithms, where a considerable number of conditions of the modelled system is at a whim of its creator. Many of the theoretical issues are overlooked or arbitrary – and therefore removed from the phenomena as such – which makes many features of genetic algorithms optional. This seems to be true for other computational methods as well (Green, 1998). Therefore, even behavioural approximations carry a huge risk of observer bias. This conundrum, approached from a different point of view, seems to be very similar to the enactivist concept of autonomy. This puzzle begs a fairly important question, asked by many (Anderson & Lebiere, 1998; Meyer & Kieras, 1997; Sun, Merrill, & Peterson, 2001, Sun, 2008b): What can we learn about cognising from computer models in terms of cognitive science? Peschl and Riegler (1999) try to offer an answer. They argue that simulations have to be taken in a certain way, that what is important and insightful “are not so much results about details, but concern conceptual knowledge which can be used as input and stimulation for both empirical and epistemological investigations” (Peschl & Riegler, 1999, pp. 15). They distinguish simulations from other empirical investigations in cognitive science and its constituents, where most of the approaches to cognition “were more or less speculative and common-sense interpretations of cognitive phenomena” (Ibid.), as progress “in empirical sciences is based on a continuous process of construction, negotiation, and adaptation to the ‘empirical data’” (Ibid.). They characterise the process of fitting theories to empirical data as a process “to reach a state of (epistemological) equilibrium in which the theory fits into the environmental dynamics, meaning that the theory—at least—predicts the environmental dynamics correctly within some margin of error” (Ibid.). However, they point out the downsides of the approach, as often “the complexity of cognitive processes and their substratum does not match the comparably poor empirical approaches and understanding of cognitive phenomena […]” (Ibid.). They feel that the more speculative areas of cognitive science open the door to simulation:

> Fortunately, the simulation method introduces a new dimension to cognitive science and, more specifically, to computational neuroscience/connectionism. Simulation models are especially interesting in the context of cognitive neuroscience, as its empirical results and theories are sometimes so rich in detail (e.g., data on the release of neurotransmitter, theories on a molecular level, etc.) that it is almost impossible to relate them to cognitive phenomena. In other words, there is an explanatory gap and a strong tension between the epistemologically inspired questions on cognition (e.g., about knowledge representation) and the empirical and highly detailed results from neuroscience. In this context the connectionist approach—in the broadest sense—plays a crucial role as mediator: it stands between the two poles of the rather speculative epistemological theories and the empirically grounded neuroscientific details and—in many cases—makes them compatible. This compatibility is achieved by the trick of focusing on the conceptual level of neural processes. By doing so, the most important characteristics and structures of neural systems, such as parallel processing, network architecture and massive connectivity, and distributed representation, are captured in a more or less simplified computational model whose dynamics can be related to and is directly relevant for epistemological and ‘cognitive’ issues. (Peschl & Riegler, 1999, pp. 15)

Peschl’s and Riegler’s faith in computer modelling is limited, and precise in that limitation. They feel it is an extremely important method, not for “the technical details of simulation which we are interested in, but rather in the conceptual implications which these models have on the problem of knowledge representation” (Ibid.). They claim that that conceptual level “can bring about both an empirically and epistemologically sound understanding of the ancient problem of representation in cognitive systems” (Peschl & Riegler, 1999, p. 16) and “guide empirical research not only on the level of technical details, but—and this seems to be even more important—on a conceptual level (e.g., concerning the assumptions/premises of a research strategy, the epistemological framework and foundations, etc.)” (Ibid.). Peschl (1991b) sums it up: “Computers are playing an important role as simulating instruments for artificial neural networks in order to achieve a deeper understanding of cognitive processes in an interdisciplinary context” (Peschl, 1991b, p. 192).

It is apparent that the term ‘computational’ comes with a great deal of conceptually ambiguous burden when it relates to being used to describe natural phenomena. This is one of the biggest reasons that some approaches have branded themselves or have been branded as anti-computational. Thinking about SEM and SMM, there is a clear distinction between the two, as SEM is inclined toward cognitivism, while SMM has been described in terms of enactivism, and enactivism rejects CtM. This begs the question whether non-computational accounts of the mind can be computationally modelled. Like wind, which does not perform computations itself, yet can be usefully computationally modelled, the mind as enactivism sees it can follow the same logic. This is why Riegler et al. (2013) question the sentiment that enactivist approaches are necessarily anti-computationalist. They are not completely clear on why that is, but a strong case is to be made for computational methods being used in regards to enactive approaches. What is clear is that enactive approaches have made considerable progress partly due to the use of computer and robotic models – and vice versa, enactive approaches definitely inspire computer modelling (Froese & Ziemke, 2009). Without presupposing that particular phenomena are computational in nature as such means that everything can be modelled and simulated – and many phenomena have been. This means that even enactive approaches can benefit from being modelled and vice versa – computer modelling can benefit from enactive philosophies. This insight bears an important gateway to thinking about whether and what we can learn about cognising from computer models. There is a peculiar similarity between the loop of benefits between enactive approaches and computer models, and the loop between knowing within cognitive science and knowing of cognitive science (COKFLONE). When considering whether the synthetic approach to natural epistemology is valuable, it has to be thought of in terms of COKFLONE. It has to have a place in that loop. This reframing of the value of computer modelling in terms of our knowledge of cognising makes producing possible answers a feasible endeavour that makes sense in terms of conceptualising cognitive science as natural epistemology. Therefore, the value of computer modelling either to knowing within cognitive science or knowing of cognitive science has to be shown, as this is where knowledge about cognising is really applied, used and where it pragmatically matters – in the circular behaviour of evolution of cognitive science.

8.1 Impact of Computer Modelling on Knowing within Cognitive Science

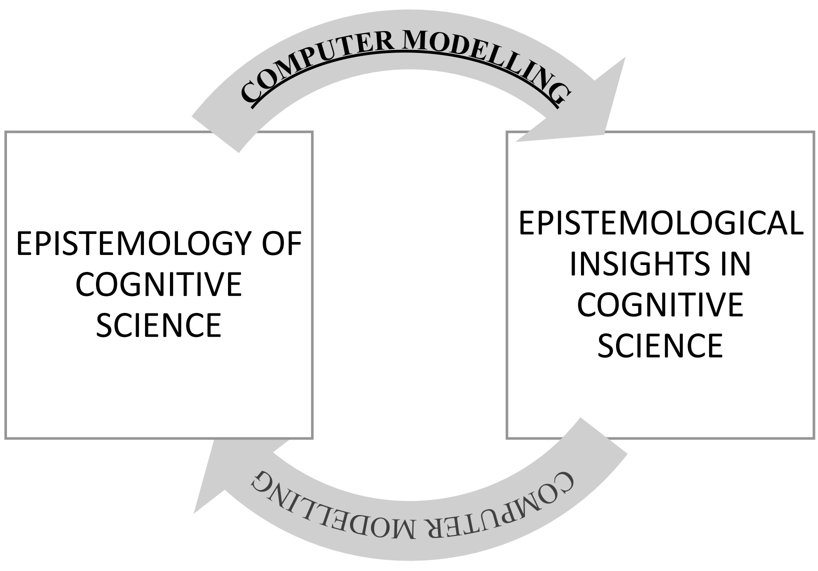

Figure 26. The impact of computer modelling on knowing within cognitive science.

In this section, a few examples where computer modelling furthered the knowledge on cognition in cognitive science will be chronologically listed. While a case can be made that positive outcomes from computer modelling research do not say anything definitive about cognition, negative do – every time a computer model fails, it is apparent how cognition does not function.

To show the broad scope of computer modelling, one example from each level of analysis, arguably pertaining to cognitive science, will be listed: “the sociological level, the psychological level, the componential level, and the physiological level” (Sun, 2008a, p. 14; see Figure 27).

level	object of analysis	type of analysis	computational model
1	inter-agent processes	social/cultural	collections of agents
2	agents	psychological	individual agents
3	intra-agent processes	componential	modular construction of agents
4	substrates	physiological	biological realization of modules

Figure 27. Hierarchy of four levels of analysis (from Sun, 2008, p. 12).

On the sociological level, which includes collective behaviour, group cognition, inter-agent processes and agent-environment interaction, most research is focused on emergent phenomena from interactions. Recently, computer models on pedestrian behaviour have been successful in determining individual behaviour from emergent properties of crowding. Moussaïd, Helbing and Theraulaz (2011) found out that “guided by visual information, namely the distance of obstructions in candidate lines of sight, pedestrians apply two simple cognitive procedures to adapt their walking speeds and directions” (Ibid., p. 6884). Their model predicts “individual trajectories and collective patterns of motion in good quantitative agreement with a large variety of empirical and experimental data”, “the emergence of self-organization phenomena, such as the spontaneous formation of unidirectional lanes or stop-and-go waves” and that “the combination of pedestrian heuristics with body collisions generates crowd turbulence at extreme densities—a phenomenon that has been observed during recent crowd disasters” (Ibid., p. 6884). They argue that their model helps give insights into spontaneous human acting in dynamic systems where there are many agents, and has already been accepted into wider accounts on so-called human crowd motion (Duives, Daamen, & Hoogendoorn, 2013).

On the psychological level, which includes individual behaviour, mental states, emotion, perception, action, concepts, development, learning and so on, computer models of child development have been successful in accounting for and giving insights into developmental processes. One of the big questions in this area is whether development and learning are distinct processes. Sun (2008a) summarises computational insights on this, which have been incorporated into child development theories and knowledge in psychology:

> Using constructive learning models also resolves the ‘paradox of development’: It was argued that if learning was done by proposing and testing hypotheses, it was not

> possible to learn anything that could not already be represented. This argument becomes irrelevant in light of constructive learning models where learning mechanisms that construct representations are separate from the representation of domain-specific knowledge. A constructive model builds representational power that it did not previously possess. Thus, computational modeling suggests that development is functionally distinct from learning. (Sun, 2008a, p. 20)

This is now the prevalent view on the issue of development versus learning (Schultz & Sirois, 2008).

On the componential level, which includes intra-agent processes or cognitive architectures, there is a number of architectures, such as CLARION, ACT-R and Soar, which boast a number of successes in various cognitive domains. CLARION has been indispensable in developing a comprehensive theory on skill learning. Sun (2008a) explains CLARION's role in understanding skill learning:

> At a theoretical level, this work explicates the interaction between implicit and explicit cognitive processes in skill learning, in contrast to the tendency of studying each type in isolation. It highlights the interaction between the two types of processes and its various effects on learning […]. At an empirical level, a model centered on such an interaction constructed based on CLARION was used to account for data in a variety of task domains: process control tasks, artificial grammar learning tasks, serial reaction time tasks, as well as some much more complex task domains […]. The model was able to explain data in these task domains, shedding light on some apparently contradictory findings (including some 20 findings once considered as casting doubt on the theoretical status of implicit learning). (Sun, 2008a, p. 20–21)

Sun argues that CLARION "helped to achieve a level of theoretical integration and explanation beyond the previous theorizing" (Sun, 2008a, p. 21).

On the physiological level, which includes biological substrates, computer modelling is mostly applied in the discipline of computational neuroscience. One enticing example of the use of neuroscientific computer modelling is understanding how blind people perceive after being implemented with technological, computational artefacts. Fine and Boynton (2015) use "a computational model of axon fibre trajectories developed using traced nerve fibre bundle trajectories extracted from fundus photographs of 55 human subjects" (Ibid., p. 4) to understand "distortions of the perceptual experience" (Ibid., p. 1). Knowledge about perception in blind people using bionic eyes gained from this research has been used in sight restoration technologies (Beyeler, Rokem, Boynton, & Fine, 2017).

These examples from each of the levels of analysis show clear contribution to knowledge on cognition and its application. In the next section, synthetic research that fuelled paradigm change and impacted knowing of cognitive science, which was necessary to advance knowing within cognitive science, will be discussed.

8.2 Impact of Computer Modelling on Knowing of Cognitive Science

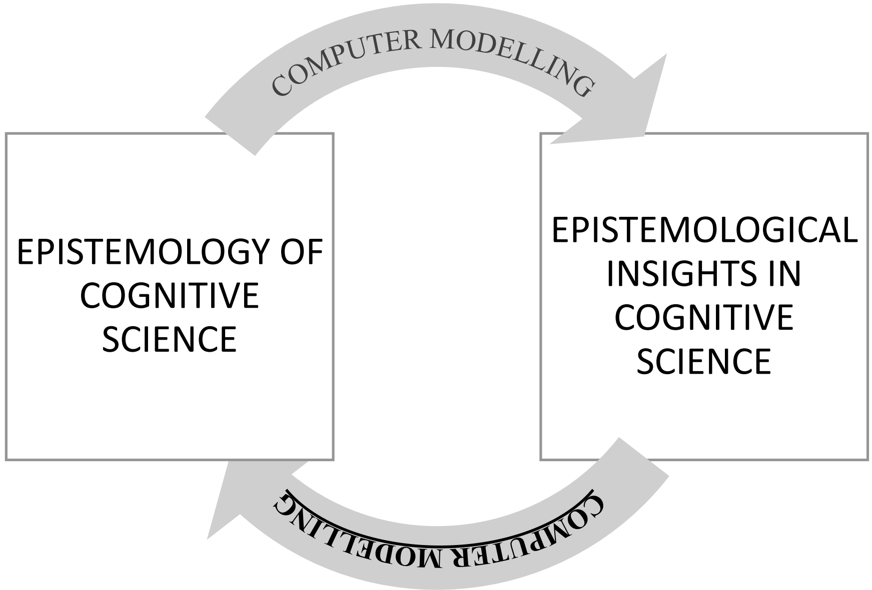

Figure 28. The impact of computer modelling on knowing of cognitive science.

In this section, paradigm changes in cognitive science and their connection to the role of computer modelling will be discussed by listing specific examples for showcasing. The paradigm changes shifted knowing of cognitive science (or vice versa, since they may be indistinguishable), and epistemological presuppositions about the mind and cognition. This lead to new knowledge in what knowing is, and new research was able to be produced on previously unsolvable phenomena, thus advancing knowing within cognitive science.

Roughly three (or, arguably, four) paradigm shifts can be discerned in cognitive science and its history, and each has had its own scientific tool, which was, tellingly, belonging to synthetic methodology (see Figure 29): cognitivism with symbolic information processing, connectivism with artificial neural networks and embodied-embedded-enactive approaches with mobile robots (Froese, 2010).

The demonstration of computer model use has to start with the first paradigm that gave birth to cognitive science. Cognitivism, which claims that cognition is computation with arbitrary symbols which represent the world, can be said to have started with one important insight. This insight can be traced back to the Pythagorean school, which tried to describe reality with abstract descriptions – e.g., they figured out "harmonious relationship between strings on an instrument which have certain simple mathematical relationships to each other (e.g. if one string is precisely double the length of another, its pitch is an octave lower)" (Ó Nualláin, 2002, p. 14). This insight that parts of the world can be usefully represented by manipulation of arbitrary symbols has found home in thinking about mind and cognition. It was due to various successes in applying this insight that the first "knowing of cognitive science" was formed. The most well-known example of this application may be modelling a chess player, the first such model written by Dietrich Prinz (who was Turing's colleague) in 1951. Progress in modelling of a computer chess player, which was supposed to represent higher cognitive functions such as abstract thinking, escalated very quickly, pinnacled by IBM's Deep Blue winning against the then greatest chess player, Kasparov, in 1996. Models like the artificial chess player signalled to cognitive science researchers that cognition is computational as well.

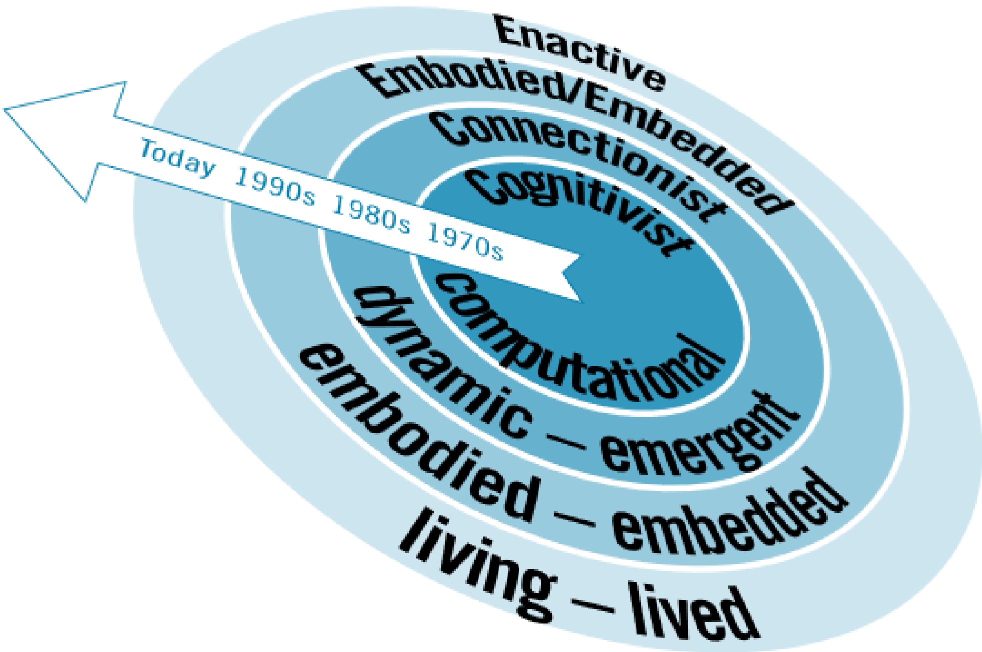

Figure 29. Illustration of one interpretation of the evolution of paradigms in cognitive science (from Froese, 2010, p. 76).

However, cognitivist computer modelling had a lot of problems with large domains, as the number of rules was too big to model top-down, making models unable to output correct behaviours. A shift in focus was needed for solving problems such as generalisation. Rumelhart and McClelland (1986) were able to predict past tense forms of English verbs, a feat previously unsolved, when they dealt with computations not in the sequential, centralised, top-down and symbolic way, but rather in a parallel, distributed and bottom-up one:

> The task is interesting because although most of the verbs in English (the regular verbs) form the past tense by adding the suffix '-ed', many of the most frequently verbs are irregular ('is' / 'was', 'come' / 'came', 'go' / 'went'). The net was first trained on a set containing a large number of irregular verbs, and later on a set of 460 verbs containing mostly regulars. The net learned the past tenses of the 460 verbs in about 200 rounds of training, and it generalized fairly well to verbs not in the training set. It even showed a good appreciation of "regularities" to be found among the irregular verbs ('send' / 'sent', 'build' / 'built'; 'blow' / 'blew', 'fly' / 'flew'). During learning, as the system was exposed to the training set containing more regular verbs, it had a tendency to overregularize, i.e., to combine both irregular and regular forms: ('break' / 'broked', instead of 'break' / 'broke'). This was corrected with more training. It is interesting to note that children are known to exhibit the same tendency to overregularize during language learning. (Garson, 2015, para. 11)

They managed to solve the task using artificial neural networks, and the design of this method itself (distributed nodes that work bottom-up) was influential in shifting the prevalent views on cognition from traditional artificial intelligence to cognition as an emergent phenomenon from a network of distributed, interconnected basic elements. However, many researchers, such as Varela and Maturana, felt that connectionism was still too focused on the brain, as connectionism classified it as the sole source of cognition, and everything, including the world – in form of representations – was crammed in it. Opposition[6] to such views thought that the connectionist account of cognition was incomplete, and consequently could not completely explain how organisms perceive and act in the world so elegantly, effectively and efficiently.

A part of the opposition and this, arguably last, monumental shift in view of what cognition is, was roboticist Brooks, who used computer modelling to explore cognition. Brooks "pointed out that the bulk of evolution had been spent getting organisms to the stage where they had useful sensory and motor systems; phenomena such as tool use, agriculture, literature, and calculus represent only the most recent few 'seconds' in the evolutionary clock" (Elman, 1998, p. 20). Brooks therefore focused his efforts on "on the hard job of building systems which have sound basic sensorimotor capacities" (Ibid.), which were until his time in the 1980s still largely unsolved. He proposed that sensorimotor abilities and direct interaction with the world, without building complex representations of it in the central unit (the brain), were crucial to understanding and unlocking the mystery of cognition. This was a fairly obscure view in mainstream cognitive science circles, although a movement was growing with similar ideas (although not all came to this conclusion and furthered the cause by using computational methods). Brooks demonstrated the value of this shift in what cognition is when he built a robot that could act intelligently in his laboratory by moving and performing simple tasks (Brooks, 1986). The relatively simple, multi-layered subsumption architecture was the key to his motto that "the world is its own best model" (Brooks, 1991, p. 15).

This was a piece in the mosaic that helped spawn a whole new way of thinking about cognition in terms of its relationship with the world, which is the main focus when knowing of cognitive science is discussed. The embodied-embedded-enactive view sees cognition as situated (that it does not deal with abstract descriptions, but rather the world here and now), embodied (that bodies non-trivially constitute cognition) and emergent (which was a large part of connectionism as well, but now emergence of cognition included the environment as well). The last part of this radical shift is that the organism enacts its world, constructs it, although it seems that computer modelling did not play an essential role in this part of the shift on knowing of cognitive science, other than the problems of computer modelling paving the way towards the shift itself.

[6] Opposition is not meant to be taken as a coherent, organised movement, but many voices from different areas of cognitive science that mainly converged in their critique.

9 Critical Analysis of the Sensory-exclusive Model (SEM) and the Sensorimotor Model (SMM)

In the previous chapter, computer modelling and its value to the endeavour of cognitive science to answer epistemological questions was examined. The value – the insights of computer modelling into what cognition is – was apparent when computer modelling was placed in COKFLONE, the loop of knowings in cognitive science. Now, the same approach will be applied to the models SEM and SMM, including the results, to discuss the implications. To be able to place the models in COKFLONE, a few separate issues have to be identified first, upon which I will base my critical evaluation.

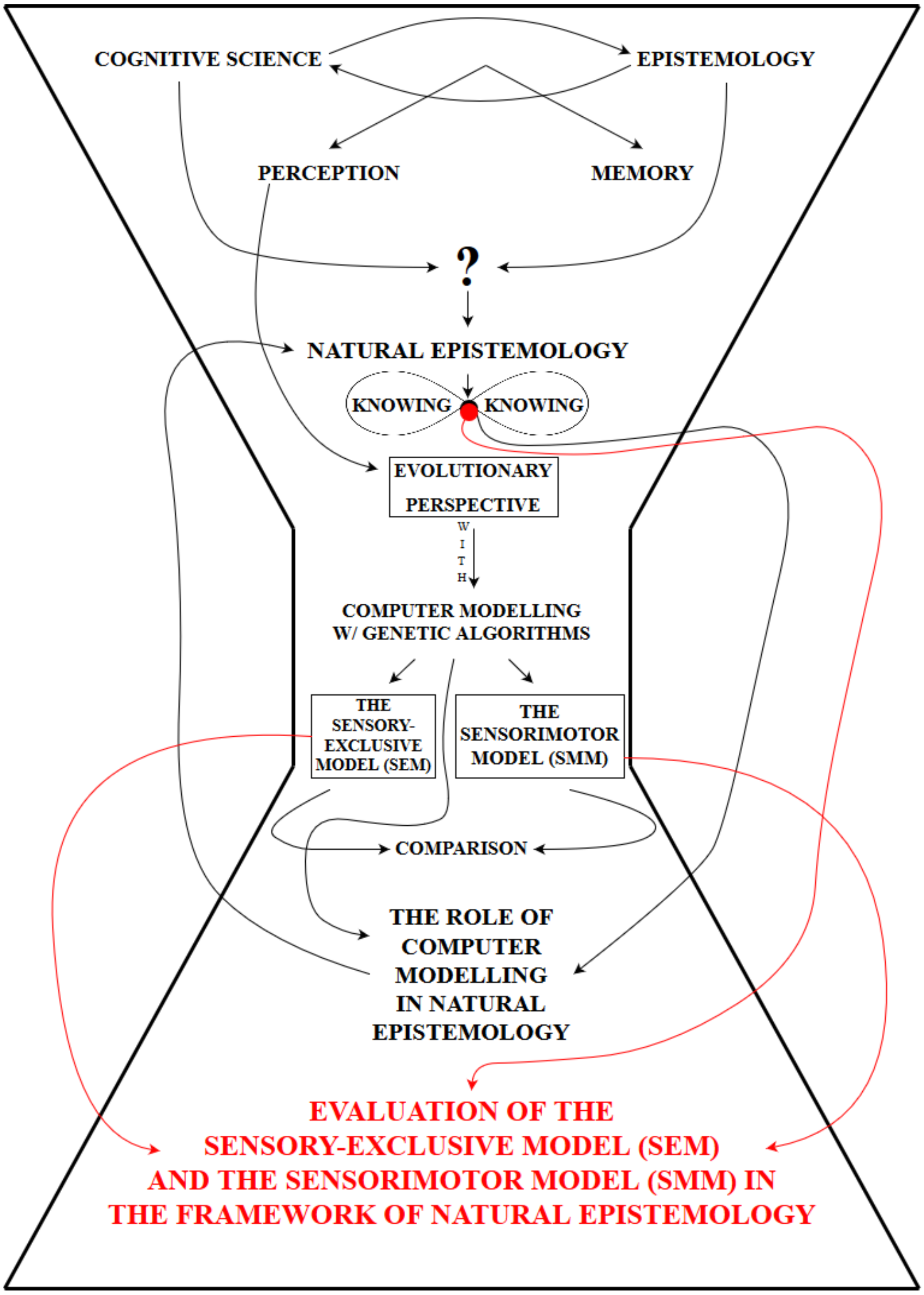

Figure 30. The position in the narrative schematic of the thesis, marked in red.

9.1 Implications of the Models' Results

Hoffman et al. (2013) accompanied the results of their model with the following words:

> It is widely assumed, by laymen and experts alike, that our perceptual systems deliver true, though not exhaustive, descriptions of the objective world. This assumption can be studied using mathematical models of evolution by natural selection. The studies thus far, using evolutionary game theory and genetic algorithms, indicate that truer perceptions are not necessarily fitter, and that a simple interface strategy can drive true perceptions to extinction. The ubiquity in nature of phenomena such as supernormal stimuli and releasing stimuli suggest that these mathematical results are not artifacts and that many, perhaps all, organisms have been shaped by natural selection not to see truth but instead to have perceptions that guide behavior that is adaptive in the relevant niche. In short, perception seems to be tuned to fitness rather than truth. Can an organism act adaptively in an environment that it does not see truly? Our GA results and those of Mitchell show that this is in fact possible. Our GAs evolve an organism that effectively forages in an artificial world that is a two-dimensional grid of squares, even though the organism itself is utterly ignorant of the two-dimensional nature of this world. (Hoffman et al., 2013, p. 10)

Hoffman and his colleagues seem to believe that their model and its results can tell us the reality of our perception. There are two possible issues with this. The first issue concerns the implication of the results and conclusions when genuinely applied to all living beings, observers included. The second issue concerns implications of the results and conclusions as they were intended by Hoffman et al.

Let's take the results from SEM seriously, believing that "genetic algorithms render a clear verdict: Natural selection discards veridical perceptions and promotes interface strategies tuned to fitness" (Hoffman et al., 2015, p. 1482). In reporting the results in such a resolute way, they are revealing a paradoxical presupposition that they are holding: that even though they themselves are living beings, organisms whose perceptions have been shaped by evolution and its mechanisms for non-isomorphic and non-veridical representation of the outside, objective world, they apparently hold a special privilege to access that same world and are able to reveal "truth" about it. They claim that they have evidence that organisms' cognitive access to the world does not match its real, true structures, yet they seem to be claiming biological realities about these same organisms as these same organisms. This highlights a blind spot on the part of the authors that has been addressed in the second-order cybernetics movement, where it is paramount to realise and address the fact that everything observed about how living beings observe is made by observers who are themselves living beings who observe, if I paraphrase Maturana's famous article title "Everything said is said by an observer" (1987).

The conundrum of such research is clear from the beginning. If one argues that their computer model sets out to discover the "truth" about whether or not we cognise the objective world around us "truthfully", one is bound to encounter serious problems with the interpretation of the results if they happen to show that we do not cognise veridically. According to ITP, simplistically stated the results from their model would mean that the authors' discovery may be an indication of their evolutionary needs to get such results, if their endeavours were to be interpreted as necessarily important for their evolutionary fitness. This is an extreme, yet possible interpretation; Peschl (1999) thinks scientific theories are something similar:

“[Scientific theories] are not understood as ‘objective descriptions’ of the environment (which is impossible anyway from an epistemological and constructivist perspective), but as strategies for successfully coping with the environment. In other words, the aspect of manipulating and predicting the environmental dynamics are the central features of scientific theories” (Peschl, 1999, p. 187). This is an interesting situation that has to be addressed further, as it also applies to SMM, although I am hesitant in claiming that the model clearly shows that natural selection leads to organisms having non-isomorphic perception.

Side-stepping the issue of the results of the models being applied to the observer and not only to the artificial organisms, taking the results as being potential evidence for how organisms truly perceive means that they answer to the epistemological question on the topic of how we cognise the world. The answer would be close to some of the claims of radical constructivism: “a. The function of cognition is adaptive, in the biological sense of the term, tending towards fit or viability; b. cognition serves the subject's organization of the experiential world, not the discovery of an objective ontological reality” (von Glasersfeld, 1995, p. 18). However, to take the results as potential evidence, the models have to be considered through a number of issues that might hinder their evidential legitimacy. Five areas worth investigating in regards to the models’ legitimacy are identified: explanatoriness, predictive power, complexity, abstractness, and viability of genetic algorithms as the method for modelling this phenomenon. Different levels of each of the areas will be discussed.

9.1.1 Explanatoriness and Predictive Power

Hoffman et al. (2015) implicitly present SEM as well as ITP more as an explanatory than a predicting tool. This is illustrated when they write about the different perceptions of dragonflies and male jewel beetles (see chapter 2 for the full quotes). Dragonflies lay eggs on very specific reflective surfaces that, at least to us, represent water, as eggs can only survive on water. Dragonflies are convinced that oil slicks and tombstones that reflect in the same specific way are arguably the same egg-laying grounds as water is. Male jewel beetles recognise females by their curvature and specific colour. Since newly introduced beer bottles boast the same attributes, male jewel beetles are convinced they are females. The theory and the model work as an explanation to these two examples. Dragonflies perceive the graves absolutely as they do the water, as the same structure, as a place to lay eggs. Male jewels absolutely think that the bottle is the same thing as the female beetle, and what is more, the bottle is the best female beetle yet. This non-veridicality is, in this example, damaging to evolutionary fitness, which leads to extinction in both cases, unless natural selection produces more successful individuals with tweaked perceptions. Either way, the perception is non-veridical. There have also been occasions where other authors used ITP for explanatory purposes of other phenomena, e.g., Kaznatcheev, Montrey and Shultz (2014) use evolutionary games to implicitly study behaviour in social phenomena like religion, explaining it with ITP, saying:

> [B]y creating agents who lack an *a priori* understanding of the world, we demonstrate that such agents can evolve a misrepresentation of reality. Furthermore, because evolution selects for adaptive behavior rather than accurate internal representations, these delusions may prove useful by encouraging a greater degree of cooperation than rationality would otherwise allow. By offering an example of how internal representations and their consequences for behavior can be studied in a game theoretic context, we hope to pave a path for understanding how and why humans deviate from objective measures of rationality. […] An interesting domain for such social interfaces

> is religion. Religion is often lauded for promoting cooperation and moral behavior […] and often criticized for disseminating incorrect and even delusional beliefs […]. When evaluating the net impact of religion, these two well-supported positions are typically placed in opposition. Our model is consistent with both of these claims, while providing an explanation of how these tendencies can emerge from the same underlying process. (Kaznatcheev et al., 2014, p. 5)

The authors therefore explain religion as a strategy for improving chances of survival in groups, yet the survival comes at a cost of misrepresenting the reality. However, explanatoriness is not predictive power. As predictive power is a major measurement for evaluating models' validity (Hand, 2014), the question is this: Can SEM and SMM predict anything? However, it is equally important to consider this: Do they have to predict anything? It is hard to make a case for a positive answer the models being able to predict anything. The relationship between mind and world in a way that is represented in the models, by Robbies' lives being reduced to foraging for soda cans and only seeing two colours, is not conceivable for interpreters to be thought of in terms of predicting specific natural phenomena that occur in the real world. The core issue itself – do we perceive the world isomorphically or non-isomorphically in general – does not offer much in the way of prediction, even if the models were constructed differently. Since access to the objective world is impossible, prediction in regards to that same world seems to be impossible as well. But the point of building such models, models like SEM and SMM, is not to predict behaviour. What such models offer, including SEM and SMM, is to provide functional understanding of what rules and conditions give rise to certain relationships between mind and world. In SEM's and SMM's case, rules that defined Robbies, the environment, the fitness function and so on resulted in non-isomorphic relationship between mind and world, a relationship that reflected not the structure of the external world, but that of Robbies' survival needs. It can be said that what these models can be viewed in predicting such relationships, but these relationships are heavily dependent on a number of factors. Two of them are important for computer models in general: complexity and abstractness. They will be discussed in the next section. Afterwards, the more difficult topic of observer-based implemented rules of nature, such as the fitness function, will be discussed.

9.1.2 Complexity and Abstractness

The task of determining the level of complexity and abstractness of a model can be approached in two ways – by looking at the three levels of organisation (Marr, 1982) as well as by looking at the results from comparable models and discern why they are similar or different. Marr delineates between three levels "which any machine carrying out an information-processing task must be understood:

1. Computational theory: What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?
2. Representation and algorithm: How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?
3. Hardware implementation: How can the representation and algorithm be realized physically?" (Marr, 1982, p. 25)

In terms of computational theory, the goal seems to be very abstract. Trying to find out whether perception is isomorphic or non-isomorphic in general is fairly high in terms of levels of analysis. The goal is not specific in terms of situations, different organisms or biological levels – the models make a sweeping generalisation of perception as such. Representing this far-reaching and complex goal algorithmically by assigning two colours to 11 quantities of resources and evolving this by simple functions of crossover and mutation shows a major simplification. This is realised in a limited simulated environment with Python, with lines of code running well below 1000, with a small number of parameters and few capabilities of the Robby class in terms of perception and action. The hardware implementation is far from being placed outside the computer, e.g., in a robotic model. This speaks further to the simplification of an extremely complex and abstract problem. It is therefore a question of whether matters of such abstractness can be modelled. Farkaš (2012) and Sun (2008b) imply that anything that can be formalised can be modelled. However, phenomena from lower levels of explanation are more easily formalised than higher, more abstract ones (lower level models are thus better at predicting).

Looking at the results of both models, what draws attention is that both have the same result in the local maxima as well as in evolving non-isomorphic perceptions, which is further underlied by very similar running times that takes them to evolve 500 generations. The only exceptions are the slopes of the fitness progression over generations, which are different for the models, but that has been explained by higher variance when the sensorimotor loop is added, which works as an optimiser (see Figures 18 and 19). There are two possible interpretations of both models producing the same results. On the one hand, this can point to the simplicity of the models, where, at this level of complexity, adding certain elements such as the sensorimotor loop might not make a vital difference. This also makes them "limited in [their] validity because of 'underdetermination' or 'equifinality' i.e. given any finite amount of evidence, there is at least one rival (contrary) model which equally fits with the data. In other words, the evidence cannot by itself determine that some one of the host of competing theories is the correct one" (Tolk, 2013, p. 62). Since there are obviously two models that fit the output data, 'equifinality' arises, which seems to occur due to the simplicity in what Tolk labels 'data'. On the other hand, simplicity only may not be a good enough explanation and the results may actually tell us something about evolution of veridical perception. To explore this further, GAs have to be thoroughly addressed.

9.1.3 Viability of Genetic Algorithms (GAs) as a Method of Natural Epistemology

Methodological merit of GAs will be discussed in two parts. First, the issue of arbitrarily set parameters and their inclination for fine-tuning, both of which are consequences of unknown biological parallels, will be discussed. Second, I will argue that parameters and functions that are set knowingly as if presenting the external world are, contrary to prevalent thought, an even bigger issue in light of GAs being used as a method for natural epistemology.

As a biologically-rooted method, GAs seem to be a double-edged sword. On the one hand, their groundedness in mechanisms of evolution offers a great modelling platform that, at least formally, possesses a number of ecological structures (e.g., the abstract mechanisms of genetic recombination and mutation). On the other hand, given their stochasticity, a number of features ingrained in these structures depend on arbitrarily set probabilities, parameters and calibrations, meaning there is usually no grounding in what we know about natural phenomena. The most problematic are parameters of the evolutionary aspect, which are a critical part of GAs when used in researching natural phenomena. This especially holds true for crossover and mutation.

Even before setting the probability, methods for crossover and mutation have to be set. This mostly encapsulates the numerous ways of a child organism to inherit its genes and the way the latter are subsequently mutated. Optimal results can be quickly gained by fine-tuning these two parameters even in more non-trivial problems. A simple solution would be that, when researching phenomena related to biology, parameters have to be set to represent them. However, in complex and dynamic systems like cognitive phenomena, the natural attributes themselves are still largely undiscovered (Bear, Connors, & Paradiso, 2016; Pinel, 2013), which pushes the translation from biology to computer models beyond the usual reduction and into the realm of bias. This is why GAs are primarily used for applied optimisation – the fine-tuning of given parameters makes sense for such usage. The use of computer modelling in cognitive science poses with an inescapable problem of how to computationally describe complex natural phenomena without being too reductive. This problem is amplified when the reduction is coupled with reductionist paradigms. However, there are only minute-solutions for problems of relating GAs to biological realities and the phenomenon of perceptual veridicality. Parameters could be set in advance and based on as rigorous research of biological counterparts as possible. Researchers should not tweak and continually run the GAs to achieve favourable results. Reporting should be transparent as possible, ensuring open source practices. Research that studies perceptual veridicality lacks such practices, which obstructs progress.

The problems of ecologically viable setting of functions and parameters to match biological findings are prevalently discussed in computer modelling of natural phenomena (Mitchell, 1999). Questioning the viability of using GAs for researching cognitive phenomena therefore seems warranted. It seems to be generally agreed that where undiscovered biological properties are concerned, it remains unclear how to proceed in order to remain as unbiased as possible (Hart & Belew, 1991). But this debate, although at the forefront of research of natural phenomena with GAs, do not cut through to the deeper issue that GAs possess. What is even more problematic may be the parameters which are not arbitrarily set (and can be therefore fine-tuned), but the parameters and functions that are knowingly set by the observer as if they reflect the external world and as if they belong to it, as opposed to the organisms themselves.

To demonstrate why the parameters and functions that are not arbitrary and are set by the observer knowingly may be even more problematic, Riegler's PacMan Syndrome has to be recalled (2001). Its relevance to the computer modelling that depends on agents is particularly important in the case of GAs of both of the models presented in the thesis. The insight that "[a]rtificial agents interact with anthropomorphically defined entities, such as 'food' and 'enemy', which make sense only to the programmer of the system" (Riegler, 2001, p. 4) can be applied to a number of algorithms and functions in GAs, but there is one function that seems to be the most important: the fitness function. What is a fitness function? A fitness function keeps track of certain behaviours of organisms and ranks the latter according to those behaviours. In the case of SEM and SMM, the agents in them possess a payoff function for the number of sodas, where the payoff varies in the shape of a bell curve (see Figure 31).

The definition of what keeps organisms alive in a GA model is dictated solely by its designer, the observer, the programmer, the researcher. How can she justify implementing any kind of fitness function without being exposed to the PacMan Syndrome? And, regardless of the PacMan Syndrome, is this fitness function related in any way to what nature does? Should the fitness function be what nature imposes upon an organism or should it be the organism's internal dynamics that manifest in its self-determination by deciding how to flourish in a given environment? Or even worse, especially in terms of complexity: is co-determination the answer to this puzzle? These are not trivial questions. The antagonism between natural and internal selection, as Riegler (2008) names it, is especially daring.

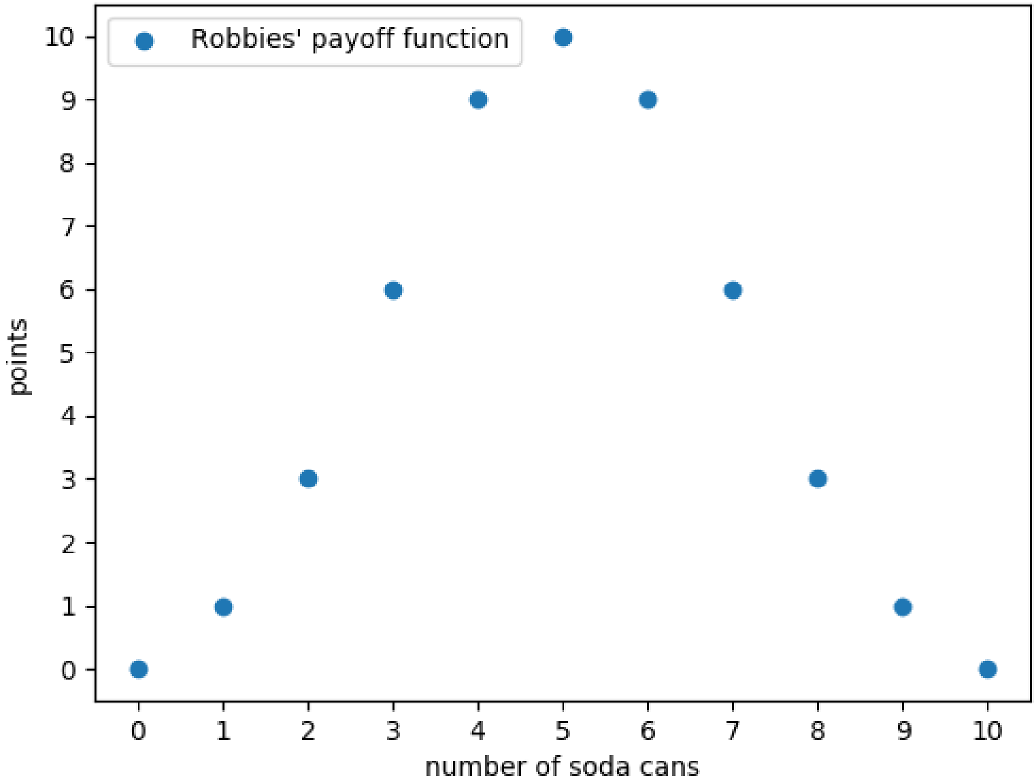

Figure 31. Robbies' payoff function – points scored for each number of soda cans – in SEM and SMM.

There are some elements of the fitness function that can be taken care of in spite of the seemingly always looming PacMan syndrome and the indeterminacy of its originator. Even if set by a designer, it has to be dynamic. Changes should occur phylogenetically as well as epigenetically. Internal goals are always built up developmentally (Oyama, 1985). The design of purpose therefore has to be a system that can be perturbed by either the modelled environment or the internal selection of the agent – ideally, both. The fitness function, wherever it is, has to be interactive and co-determined. By making it static, it is fully and, most importantly, predictably in the hands of the designer. If the fitness function itself is dynamic, influenced by environmental and organismic perturbations, even when the environment and the organism are again designed by a researcher, the fitness function no longer has a predictable role and carries much more weight in functional experiments. The biggest presupposition, embedded in a model, is, by making it dynamic, ruled by rules that result in complex interactions, which means the fitness function is much more amenable to being something that is at least on the surface slightly removed from its designer. Concretely, this would mean that its designer would not be able to predict the goals of the organism, as the initial specification would be changed by evolution and ontogeny. The initial goals and purpose would not be the final goals and purpose, making the fitness function ever so slightly disassociated from its designer. The distinction of where the fitness function belongs to, to the workings of nature or to the internality of the organism, is blurred. This limbo that the fitness function finds itself in when perturbed phylogenetically and epigenetically from rules set in the environment as well as in the organism is nowhere near the answer to the wider problem of the PacMan Syndrome and the originator of the fitness function, but it definitely makes it less distinguishable from its designer's role of the creator, as she is no longer all-knowing in imposing purpose on the organism.

The lack of answers regarding computer modelling and GAs, especially the problem of the PacMan Syndrome and the fitness function originator, may seem disheartening. Addressing the topic seems to only provide more uncertainty and push the merit of synthetic methodology for natural epistemology towards fruitlessness more and more. But are these objections the end-all for SEM's and SMM's value to cognitive science as natural epistemology? Can really nothing be learned from them? It is apparent that computer modelling generally is useful for advancing cognitive science as a discipline, as shown in chapter 8. The considerations in the present section apply to each and every endeavour listed there as well. Apparently, that has not stopped computer modelling from having serious, direct and useful implications for advancement of cognitive science as natural epistemology. In the next section, the insights and potential impact from SEM and SMM will be discussed, as it is only now that the implications of the models' results can be fully discerned.

9.2 Insights from and Potential Impact of the Models in the Framework of Natural Epistemology

Maturana's, Varela's and Hoffman's analogical stories from chapter 7 on the discrepancy between the external world and the pilot's, the navigator's and the computer user's interfaces that were operated with for successfully acting in the world have a subtle and possibly unintended percept. The percept is surprisingly insightful specifically in terms of computer modelling of epistemological issues. When Maturana's pilot described what his concept of the world was, namely the cockpit instruments, which bear little resemblance to the external world, yet ensure the pilot's successful landing of the plane, there was an outside observer, his wife, who could discern the contours of the "real" world in the analogy and possibly compare it to the cockpit instruments. In epistemology, there is no such thing. How can perceptions, which are analogous to the cockpit instruments in the pilot story, be compared with the "real" world, when there is no observer to access it?

Regardless of the numerous issues that computer modelling faces, SEM and SMM seem to be the only tools that enable that, albeit in a modelled, hypothesised situation. SEM and SMM have a designer who has access to the external modelled world and can therefore compare it to the perceptions of modelled organisms. All the objections like the PacMan Syndrome and the fitness function originator still apply, yet computer modelling seems to be the only approach to science that can honestly know the external world – only the external world is a modelled one. This is where the fact that everything is designed by a programmer and therefore the design reflects the programmer's rather than a model's innate concepts is not a bug, but a feature of computer modelling. It is a strength rather than a weakness. The external world of SMM is known to me in its entirety, which is why I was able to compare it with the organism's internal world. The same goes for Hoffman and his team. I believe this to be a part of a key insight. It means that if the world is structured in a particular way, manifested in a model through its designer's presuppositions, the results are valid for the world that exists in that particular way. This single element – knowing the entirety of the model's world – makes computer modelling useful especially for natural epistemology, which makes SEM and SMM useful for natural epistemology.

SEM and SMM can now be addressed further in terms of their position in the framework of natural epistemology, their role, hopefully, more understood. Two questions will be addressed: 1) Do SEM and SMM tell us anything about cognition? 2) If they do, is the approach of making models that vary in presuppositions about the mind, which are independent from hypotheses, and comparing them, which was undertaken in this thesis, a useful approach – for investigating the original phenomenon as well as for testing different paradigms' proclivities towards the investigated phenomenon?

I will try to address the questions in reverse. SEM and SMM were made in such a way that the only difference between them are presuppositions that are associated with different paradigms in cognitive science. Chapter 6 and parts of this chapter show how complexity overall undergoes no significant change (even though there was a change in complexity of one the perceptual DNA), which implementionally directly manifests in the very similar running times when Robbies are evolved 500 times. Order of the function therefore stays the same, which is of major importance when talking about comparability (Sipser, 1997). The major difference between the models is the inclusion of the sensorimotor loop, which makes perception and action inseparable in SMM, as well as actions having direct influence on what and how much of the world Robbies perceive. The fact that Robbies cannot perceive everything around them makes actions of where to turn to perceive really significant in their behaviour. It was important that the sensorimotor loop was not connected to isomorphic or non-isomorphic perception as such, as it is interesting to observe whether hypotheses-independent presuppositions can influence the results. SEM can therefore be plainly classified as cognitivist and SMM as enactivist, although there are problematic implications of such classification, which were discussed in chapter 6. The most important factor that can help address question 2 is whether the models are complex enough that such different presuppositions could potentially lead to different results. In other words – are these presuppositions strong enough to nudge the fitness function, the most important element in determining Robbies' perception, in one way or another regarding perceptual veridicality? What can be seen from the results is that the sensorimotor loop had a profound and novel effect on Robbies' evolution. It served as a kind of an optimiser, producing fitter Robbies from the start and therefore bootstrapping their fast evolution to a local maximum score. It therefore made a difference, which is interesting for exploration of different presuppositions on behaviour, but it did not make a difference regarding the final result in perceptual veridicality. I think that the models are not constructed in a way where fitness function-independent presuppositions about cognition are able to make a difference in resulting perceptual veridicality as long as the models' complexity stays the same. This means that varying presuppositions of different paradigms of cognitive science would always result in non-isomorphic perception due to the all-determining fitness function. However, this does not mean that comparing more complex models with this approach is not useful, which I believe it would be, and what is more, it is interesting to observe how presuppositions influence development, even though the final result is the same, as knowledge about development is vital as well. Since the influence of the sensorimotor loop in SMM was not expected, this conclusion is much stronger for it – the sensorimotor loop offered Robbies a wider range of possible behaviours that were meaningful in the sense that each action and each perception mattered so much more, and this is despite SMM's Robbies perceiving their environment in a more limited way than SEM's Robbies, from 5 spaces to only 2 at a time. The comparison of the models may also hint at the importance of the fitness function as such, along with the possibility that it is not their simplicity and the role of the fitness function in their design that produce the result, but that the results may simply be legitimate in what they report.

The most straightforward answer to the first question, yet no less true and insightful, comes in the form of a conditional. If the mind is analogous to a computer (therefore adhering to the term "computational" in the first sense as described by Riegler, Stewart and Ziemke (2013), see chapter 8) or if it can be sufficiently formally described to become a useful tool for thinking about the nature of things (adhering to the term "computational" in the second sense), and SEM or SMM accomplished to successfully model one or the other, then the models do tell us something about cognition or at least guide our thoughts on it. It has been established that even if the mind is not analogous to a computer, the models have proved to be useful either in representing the nature of things to be useful in guiding our actions as human being or useful for thinking about cognition to advance our knowing. This hints that the only way to answer the question in any satisfying, but non-trivial form is to hypothesise about the potential impact of SEM and SMM on knowing within cognitive science and knowing of cognitive science, as this seems to be where the verified, practical benefits come from when talking about advancing our knowing on cognition.

9.2.1 Potential Impact of the Models on Knowing within Cognitive Science

To discern the potential impact of the models on knowing within cognitive science, it may be useful to repeat the fundamental questions that Hoffman posed before he proposed ITP and started to build research around it with his colleagues:

> First, is the vocabulary of our perceptions isomorphic to aspects of objective reality so that our language of perceptions could, in principle, describe the objective truth? Second, if so, do our perceptual systems, using that vocabulary, in fact succeed in describing the true state of affairs in the world? (Hoffman et al., 2015, p. 1482)

SEM addresses a part of these two questions, so it is important to focus on what it tries to answer. Hoffman el al. (2015) develop their GA model to answer the question whether isomorphic perceptions ever evolve alongside non-isomorphic perceptions. They came to this question after finding out that when they used evolutionary games to study their two initial questions "that veridical perceptions fare poorly against interface perceptions when both are on the same playing field" (Hoffman et al., 2015, p. 1487). They figured out there was a "prior question to be asked: Will veridical perceptions even get on the playing field? Or are they so unfit that evolution is likely to pass them over completely?" (Ibid.) Both, SEM and SMM, seem to give a straight answer to that. However, this is exactly why the fitness function of the models is so problematic – the answer entirely depends on it. Were the payoff constructed not non-monotonically, i.e., (0,1,3,6,9,10,9,6,3,1,0), but rather monotonically, i.e., (1,2,3,4,5,6,7,8,9,10), the outcome of Robbies' evolution would be isomorphic perceptions. So what the models really tell us is something different, yet also important: that perceptions strongly conform to the organisms' internal dynamics rather than perceptions conforming to the external pressures. And this is true and important even if the internal dynamics are isomorphic to external perturbations – the direct influence is still (internal dynamics→perceptions) rather than (external perturbations→perceptions). This is an important insight that is not particularly novel, as it has been expressed by mainstream biology, as presented in chapter 2. However, intuition of mainstream biology took precedence in it persisting that perceptions are, even though they depend on organisms' internal needs, still veridical. This intuition may come from a functional blind spot as to how non-isomorphic perception could form, which is where SEM and SMM shine and can really offer some, though simple, yet valid explanatory answers as to how non-isomorphism in perceptions can work and be evolutionary beneficial.

I believe this is the extent of the impact that can be honestly argued for – looking directly at the question that spawned SEM and see what impact it can have with SEM's and SMM's answers. This is also because, regardless from the role of the fitness function determining the results, the hypothesis as it is can never be truly confirmed.

Apart from giving a functional understanding of fit non-isomorphic perceptions that may offer a way to abandon intuitions on veridicality, SMM makes a case for the importance of sensorimotor loops in interacting with the environment, as it directly offers more potential behaviours that are at the same time more meaningful in terms of forming a coherent account of perception and action. The fact that SMM resulted in faster achievement of the local maximum, therefore having a particular usefulness, as well as being taunted as very important in the first case, speaks to an unpredicted consilience. This means that having two sources claiming the importance of the sensorimotor loop, the comprehensive biological approach in enactivism and in SMM's computational manifestation, with the latter not being augmented with the sensorimotor loop for optimisation purposes, bolsters the credibility as it emerges as a consequence of research in two different areas. There is a possible conclusion that can be reached regarding the cognitive and enactive paradigm. As the sensorimotor loop can be identified as a useful mechanism when comparing SEM and SMM, it seems that at least this is where enactivism holds credence in its criticism of cognitivism and presuppositions regarding sensorimotor contingencies. Such conclusion comes close to how SEM and SMM can potentially impact knowing of cognitive science, which will be discussed in the next section. This is also a good example that loosely delineates knowing within cognitive science and knowing of cognitive science – the results of the models can impact the knowing within cognitive science, while the changing presuppositions is what can influence the knowing of cognitive science. It also gives another argument why the approach of comparing models with different presuppositions is a good approach, especially in terms of natural epistemology.

9.2.2 Potential Impact of the Models on Knowing of Cognitive Science

Hypothesising about the potential impact of SEM and SMM on knowing of cognitive science is more difficult than on knowing within cognitive science. It seems that the latter is more predictable, as it usually follows directly from the results of a given research endeavour. The more general shift in presuppositions that may occur in cognitive science as natural epistemology is less predictable, as it cannot be pinned down to where it originates from. It is not meant that it is hard to point at one single research that causes a shift – it is usually a gradually building movement – but in terms of SEM and SMM, it is difficult to predict what part of the endeavour can potentially influence what presupposition, intuition, attitude, etc. in cognitive science. Where can a shift occur? One possible source of influence in regards to paradigm frameworks has already been mentioned: the sensorimotor loop. However, it seems improper to present the sensorimotor loop as something that may cause a shift as the importance of the sensorimotor loop is not an obscure view. SMM may in that regard further cement the sensorimotor loop as a major constituent of cognition, but I do not believe that there is a shift that may occur because of it. There is another source that may be impactful: the fitness function. This is less firm as this thesis explores it in a rather abstract and simplified way, but what the thesis shows is the influence of a particularly designed fitness function on perception in a model with certain presuppositions. Since the presuppositions do not seem to have a direct influence on the evolved perception the fitness function seems to be the one guiding it, as opposed the certain paradigm specificities or the external world as modelled in SEM and SMM. What the thesis also showed was that the fitness function is not obvious and is hard to pin down – emphasising the fact that we do not know what the fitness function is, but if we presuppose to know what it is, we get interesting results. Even when pinned down, it is pinned down by a designer and not self-determined by an organism. What the analysis of the fitness function points to is the notion of the importance of a non-trivial autonomy (see chapter 6, section 6.1.1.1, "Autonomy"). This is a specific kind of autonomy, as was presented in reference to Froese et al.'s analysis of different autonomies (Froese, 2007), that has still not taken roots in agent-based computer modelling and artificial intelligence. Non-trivial autonomy is when agents are self-determining, where they self-produce their purpose and goals as a result of self-production. Riegler (2001) identified the problem of inability of computer modelling to produce non-trivial autonomy in agents as the PacMan Syndrome. The problem is known. However, SEM and SMM do not seem to directly offer solutions. So the possibility of shifts in these areas is unpredictable as well.

The general notion of ITP that evolution strives towards organisms perceiving non-isomorphically, or more correctly, that organisms' perception is fitness-based, is closely related to what was discussed on the fitness function. The PacMan Syndrome directly leads to the agents' self-determination through the fitness function, which is what dictates perceptual veridicality. This insight is similar to what the paradigm of radical constructivism notes in its concept of viability: "Briefly stated, concepts, theories, and cognitive structures in general, are viable and survive as long they serve the purposes to which they are put, as long as they more or less reliably get us what we want" (von Glasersfeld, 1981, p. 91). SEM and SMM may therefore offer a gateway to radical constructivism, however, radical constructivism has been around of decades. Taking the results seriously and applying them to researchers, who are organisms as well – and therefore taking the route of second-order cybernetics – may shift a view on scientific research itself. Peschl (1999) proposes that scientific theories are strategies to survive in the world rather than a tool to objectively describe the world as it is, which may be exactly what the shift might be, and in this case, it can be said to be a shift, as this does not seem to be a common idea in any sense of the word.

However, everything listed so far references something known and not something new. All the possible shifts discussed have already happened, albeit to different degrees. This is not so problematic, as it has to be acknowledged that predicting something like a paradigm shift is near prophetic. The possible impact of SEM and SMM on knowing of cognitive science, the impact that would be wholly new and its own, may be unimaginable. The outcome may also be that no impact happens at all.

10 Discussion

The thesis explored the role that computer modelling plays as a method in the framework of cognitive science as natural epistemology, the investigation of epistemological questions with scientific methods. This endeavour complexified at each level that it was analysed on.

First, natural epistemology was defined through elaborate designing of a network that exists between cognitive science and epistemology. The overlap between the two was identified as natural epistemology, defined as the study of epistemological questions with the use of natural scientific methods. A loose model was proposed related to how progress works in cognitive science as natural epistemology, where the key element was the loop between the knowing within cognitive science – epistemological insights in cognitive science – and knowing of cognitive science – epistemology of cognitive science. The loop signified how progress in this endeavour was made: while researching knowing within cognitive science, which is influenced by epistemology of cognitive science, there is only so far that a certain paradigm can go, as historically noted. When research hits a wall, a shift in knowing of cognitive science is necessary to get to new insights in cognitive science.

Second, perception was presented as a key source of knowledge (as proposed by empiricists) when one of the main epistemological issues and a topic of the thesis, the relationship between mind and world, was discussed. One of the foremost areas of interest in relation to perception was described, the evolutionary perspective on perception. The prevalent views were demonstrated to be fairly one-sided in favour of perception and, more widely, cognition, faithfully representing the external world in an internal replica. However, many of the proponents of these views were found to be contradictory in their statements that perception exists to serve organisms for survival according to their unique needs and not to represent the external world as it is. It was therefore suggested that the researchers' intuition that there is an isomorphic relationship between mind and world – a relationship where the mind preserves the structures of the physical make-up of the external, causal system – may not be as well founded as it is generally thought to be. Various researchers in the intersection between philosophy and artificial intelligence believe that using computer modelling may help resolve such disputes. One such model was identified in a genetic algorithm model by Hoffman et al. (2015), who used it to validate Hoffman's interface theory of perception.

Third, the interface theory of perception was presented, whose focal point is that organisms' perception works as an interface between the external world and the mind, showing the world as to guide evolutionarily beneficial behaviour and not to represent the world as it is. The authors of the theory claimed that this interface presented the world to organisms non-isomorphically. This was studied with a computer model, which used a genetic algorithm to simulate natural selection. The model was presented and reproduced successfully, the results showing that what evolves in artificial organisms is non-isomorphical perception of the world. The model, dubbed as the sensory-exclusive model, was analysed to discern what (epistemological) presuppositions it possessed, and it was found that they were analogous to presuppositions of cognitivism, namely that cognition works as an information-processing machine that produces perceptions without the need for motor behaviour.

Fourth, another cognitive science paradigm was presented alongside cognitivism – enactivism. Enactivism was thoroughly examined in its claims and presuppositions to gauge how the sensory-exclusive model could be modified to carry different presuppositions in order to examine how they influence the results. Two questions were key: whether the enactivist elements at hand could be computationally described and whether it was sensible to include them in the modified model. The sensorimotor loop was found to be the perfect candidate, as it has had a comprehensive history of being computationally described and it could had been included in the modified model without losing the ability to have it compared to the sensory-exclusive model. The sensorimotor loop also directly opposed the sensory-exclusive model's author's claim that motor movement is not needed for visual experience. The modified model was therefore dubbed as the sensorimotor model. Its results were, in abstract, the same as the results from the sensory-exclusive model, as both produced non-isomorphic perception when evolving the models' agents. The major specific difference was that the sensorimotor loop worked as an optimiser to bootstrap evolution and achieve the local maximum much faster. This was identified to be due to sensorimotor model's agents having more meaningful choices because of the sensorimotor loop, which was manifested in a much higher variance of the starting fitness of the agents.

Fifth, the knowledge gained from the hands-on computer modelling of the two models was applied to critically assess computer modelling in general in the framework of cognitive science as natural epistemology. Computer modelling was placed in the loop of knowings to determine its value. Historically computer modelling had major influence in advancing the knowing within cognitive science and therefore providing numerous epistemological insights as well as serving as a key actor in shifting epistemology of cognitive science by changing the knowing of cognitive science. This was shown through describing the role of this synthetic method when new paradigms emerged in cognitive science, starting with cognitivism through connectionism to embodied approaches. However, a number of issues was identified and addressed with using computer modelling for epistemological questions, the biggest one being what Riegler (2001) dubbed as the PacMan Syndrome, the problem that agents were always implemented in a way where concepts, important for their survival and subsequent behaviour, were imposed onto them by their creators, as opposed to being self-determinate as biological organisms are.

Sixth, a similarly rigorous analysis was applied to the implemented sensory-exclusive model and sensorimotor model, with the advantage of being much more specific in examining how different parts of the models brought value to them as examples of natural epistemology research. Their explanatoriness and predictive power, complexity and abstractness, and the viability of using genetic algorithms as a method was scrutinised. In terms of explanatoriness and predictive power, the models were found to have explanatory value especially in terms of the interface theory of perception, while predictive power was more problematic. It was argued that the models were not meant for predicting natural phenomena but rather serving as a tool for functional understanding of how perception could work non-isomorphically. This was labelled as especially useful to shift intuitions of mainstream biology in their claim of isomorphism between mind and world. In terms of complexity and abstractness, the models were found to be fairly simple and highly abstract, therefore being much harder to use in terms of explanatoriness and predictive power. It was also argued that there is a high possibility of results being the same regardless of different presuppositions because of the models' simplicity. In terms of evaluating genetic algorithms as a method for natural epistemology, two categories of genetic algorithms' elements were problematised: the arbitrary settings that have no biological parallels and the settings where biological knowledge is presumed rather than critically assessed. The first was found less problematic than the second. The problem of the first category is that due to the arbitrary settings of the model, fine-tuning can lead to hypothesis-tailored results. The problem of the second category is the fitness function, where it is imposed upon the organisms how they are evaluated to be fit for survival, which means that it is imposed onto them what their needs and goals are. This was found to be an instantiation of the PacMan Syndrome. Although the problem was deeply addressed, no firm solutions were found. However, it was argued that one way of distancing the designer from the fitness function was to make it dynamic rather than static, changing phylogenetically as well as ontogenetically, and thus resulting in an unpredictable fitness function. This unpredictability in relation to the designer makes it less fundamental in influencing the needs of the organism, as the fitness function evolves by being perturbed by external and internal forces. In that way, the designer cannot predict what the fitness function can result in, making her detached in her ability to design organisms' needs that firmly influence its behaviour in a predictably intended way. At the end of this part, the potential impact of the sensory-exclusive model and the sensorimotor model on knowing within and knowing of cognitive science was speculated on. Regarding knowing within cognitive science, where models' results are what is important as opposed to comparison of the models in regards to their presuppositions, it was argued that the non-isomorphic perceptual outcome of the model might affect the intuition on isomorphism between mind and world in mainstream science of perception. It was also argued that the fact that the sensorimotor loop in the sensorimotor model worked as an optimiser might further solidify the idea of importance of sensorimotor contingencies in cognition, but this was identified as being dangerously close to influencing the knowing of cognitive science (knowing within cognitive science and knowing of cognitive science cannot be clearly and fully separated). Regarding knowing of cognitive science, it was speculated that the models may be part of a number of shifts, albeit these shifts had already taken place, rendering the potential impact in these areas idle. It was concluded that it seems impossible to predict what a potential influence on epistemology of cognitive science of the models may be that would be wholly new and their own. It was also argued that the approach of constructing comparable computer models with different presuppositions seems useful, but the models have to be sufficiently complex so that changing presuppositions can, in theory, influence the final result – not influencing does not necessarily mean lack of complexity, but rather a different insight on the investigated phenomenon.

10.1 Possible Improvements in Future Work

The biggest possible improvement in terms of computer modelling is definitely addressing the PacMan Syndrome, manifested as the fitness function in the sensory-exclusive model and the sensorimotor model. The possibility to distance the fitness function's designer from it is by making it dynamic rather than static (research in intrinsic motivation in artificial intelligence (Baldassarre et al., 2014) emphasises this) and developing it phylogenetically (from one generation to another) as well as ontogenetically (during one lifespan). This is not a solution by any means, but it is a conceptually interesting approach to lower the level of knowing of the designer, to minimise intentional imposing of needs and goals for specific agendas and to therefore make the fitness function uncertain in its manifestation.

An intriguing notion would also be to make a transition from simulations to embodying the models into a robot. This is a potentially significant transition as robotic models were essential in shifting epistemology of cognitive science in order for a new, embodied paradigm to emerge. "Brooksifying" (after roboticist Brooks, whose robots were crucial in advancing the knowing of cognitive science to the embodied-embedded paradigm, as well as anti-representational ideas on the relationship between mind and world) the models would need a careful consideration of how to define the needs in the real world and the function that would translate input into different perceptual veridicalities. If SMM were to be implemented in this way, it would be interesting to discuss including other enactivist characteristics alongside the sensorimotor loop.

The approach of comparing models' result that possess different presuppositions has been potentially deemed valid only if the presuppositions were to have the potential to influence the final result. To further investigate this approach, the models would have to be made more complex and studied to understand how presuppositions and outcomes potentially connect.

Regarding the framework of natural epistemology, the loose concept of the knowledge loop between epistemology of cognitive science and epistemological insights in cognitive science can be further elaborated and investigated in greater depth to see if it is a strong enough concept for further use. This would be an entirely theoretical endeavour, which would help build on the ideas of this thesis regarding cognitive science as natural epistemology.

10.2 Conclusion

The thesis has no one takeaway. I am not certain whether this is productive or not, but it seems to be the consequence of the toll that a thesis with such a wide net of ideas, concepts, areas, disciplines, approaches and levels, which have to integrate, demands. If nothing else, the thesis underlines that a computer model is never just a computer model, as "computer modelling of the influence of natural selection on perceptual veridicality" has shown that the burden of computer modelling comes with more baggage than commonly bargained for, and may as such tell more about the modeller than about the modelled. Methodological and conceptual dead ends, pinnacled by the PacMan Syndrome, seem to offer little room for solutions. Hopefully, the loop of knowing in cognitive science proves to be an isomorphic description of what happens in cognitive science and that the PacMan Syndrome is just another wall that necessitates a shift in the knowing of cognitive science for further flourishing of such computational endeavours. Until then, to paraphrase Dennett (see chapter 1, section 1.1.5, "Natural Epistemology"), the only way forward is to try to barrage the wall by continued thinking with bare brains as well as bare hands.

11 References

Adolphs, R. (2015). The unsolved problems of neuroscience. *Trends in Cognitive Sciences*, *19*(4), 173–175. doi:10.1016/j.tics.2015.01.007

Alston, W. (1999). Perceptual Knowledge. In J. Greco & E. Sosa (Eds.), *The Blackwell Guide to Epistemology* (pp. 223–242). Oxford: Blackwell.

Amundson, R. (1983). E. C. Tolman and the Intervening Variable: A Study in the Epistemological History of Psychology. *Philosophy of Science*, *50*(2), 268–282. doi:10.1086/289109

Anderson, J., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Baldassarre, G., Stafford, T., Mirolli, M., Redgrave, P., Ryan, R. M., & Barto, A. (2014). Intrinsic motivations and open-ended development in animals, humans, and robots: an overview. *Frontiers in Psychology*, *5*, 985. doi:10.3389/fpsyg.2014.00985

Barandiaran, X. E. (2017). Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency. *Topoi*, *36*(3), 409–430. doi:10.1007/s11245-016-9365-4

Bateson, G. (1979). *Mind and Nature. A Necessary Unity*. New York, NY: E. P. Dutton.

Bear, M. F., Connors, B. W., & Paradiso, M. A. (2016). *Neuroscience: exploring the brain*. Philadelphia: Wolters Kluwer.

Berkeley, G. (1710/1982). *A treatise concerning the principles of human knowledge*. Hackett. (Original work published in 1710)

Bermúdez, J. L. (2014). *Cognitive Science: An Introduction to the Study of Mind* (2nd ed.). Cambridge: Cambridge University Press.

Beyeler, M., Rokem, A., Boyton, G. M., & Fine, I. (2017). Learning to see again: biological constraints on cortical plasticity and the implications for sight restoration technologies. *Journal of Neural Engineering*, *14*(5), 051003. doi:10.1088/1741-2552/aa795e

Blackmore, S. J., Brelstaff, G., Nelson, K., & Troscianko, T. (1995). Is the richness of our visual world an illusion? Trans-saccadic memory for complex scenes. *Perception*, *24*(9), 1075–1081. doi:10.1068/p241075

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. doi:10.1037/a0026450

Brooks, R. (1986). A robust layered control system for a mobile robot. *Journal of Robotics & Automation*, *2*, 14–23. doi:10.1109/jra.1986.1087032

Brooks, R. A. (1991). Intelligence without Representation. *Artificial Intelligence*, *47*, 139–159. doi:10.1016/0004-3702(91)90053-m

Chalmers, D. J. (1995). On implementing a computation. *Minds and Machines*, *4*, 391–402. doi:10.1007/bf00974166

Chrisley, R. J., & Parthemore, P. (2007). Synthetic Phenomenology: Exploiting Embodiment to Specify the Non-Conceptual Content of Visual Experience. *Journal of Consciousness Studies*, *14*(7), 44–58.

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.

Ciulla, J. B. (2011). Handmaiden and queen: What philosophers find in the question: "What is a leader?" In M. Harvey & R. E. Riggio (Eds.), *Leadership studies: The dialogue of disciplines* (pp. 54–65). Cheltenham, UK: Edward Elgar Publishing Limited.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–253. doi:10.1017/S0140525X12000477

Clark-Foos, A., Brewer, G., & Marsh, R. L. (2015). Judging the reality of others' memories. *Memory*, *23*(3), 427–436. doi:10.1080/09658211.2014.893364

Cognitive Science At Simon Fraser University : FAQ. (n.d.). Retrieved August 21, 2018, from https://bcs.mit.edu/research/cognitive-science

Cognitive Science. (n.d.). Retrieved August 21, 2018, from https://bcs.mit.edu/research/cognitive-science

Cognize. (2018). In *Merriam-Webster.com*, Retrieved July 22, 2018, from https://www.merriam-webster.com/dictionary/cognize.

Daugman, J. G. (1993). Brain metaphor and brain theory. In E. L. Schwartz (Ed.), *Computational neuroscience* (pp. 9–18). Cambridge, MA: MIT Press.

Dennett, D. (1996). *Kinds of Minds*. New York: Basic Books.

Di Paolo, E. A., Rohde, M., & De Jaegher, H. (2007). Horizons for the enactive mind: Values, social interaction, and play. In J. Stewart, O. Gapenne & E. Di Paolo (Eds.), *Enaction: towards a new paradigm for cognitive science* (pp. 33–87). Cambridge, MA: MIT Press.

Di Paolo, E., Buhrmann, T., & Barandiaran, X. E. (2017). *Sensorimotor life: an enactive proposal*. Oxford University Press.

Dino, Ray. (2015). A genetic algorithm to create an evolving cleaning robot. GitHub repository, https://github.com/ray-dino/robby-genetic-algorithm.

Duives, D. C., Daamen, W., & Hoogendoorn, S. P. (2013). State-of-the-art crowd motion simulation models. *Transportation Research Part C: Emerging Technologies*, *37*, 193–209. doi:10.1016/j.trc.2013.02.005

Eiben, A. E., & Smith, J. E. (2003). *Introduction to Evolutionary Computing*. Berlin: Springer-Verlag Berlin Heidelberg.

Elman, J. L. (1998). Connectionism, artificial life, and dynamical systems: new approaches to old questions. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (pp. 488–505). Hoboken, NJ: Blackwell Science.

Fabricius, W. V. (1983). Piaget's theory of knowledge: Its philosophical context. *Human Development*, *26*(6), 325–334. doi:10.1159/000272894

Falk, J. H. (2018). *Born to Choose: Evolution, Self, and Well-Being*. Abingdon: Routledge.

Farkaš, I. (2012). Indispensability of Computational Modeling in Cognitive Science. *Journal of Cognitive Science*, *13*(4), 401–429. doi:10.17791/jcs.2012.13.4.401

Fields, C., Hoffman, D. D., Prakash, C., & Prentner, R. (2017). Eigenforms, Interfaces and Holographic Encoding: Toward an Evolutionary Account of Objects and Spacetime. *Constructivist Foundations*, *12*(3), 265–291.

Fine, I., & Boynton, G. M. (2015). Pulse trains to percepts: the challenge of creating a perceptually intelligible world with sight recovery technologies. *Philosophical Transactions of the Royal Society B, 370*, 20140208. doi:10.1098/rstb.2014.0208

Friedenberg, J., & Silverman, G. (2016). *Cognitive science: An introduction to the study of mind*. Thousand Oaks: SAGE Publications.

Froese, T. (2007). On the Role of AI in the Ongoing Paradigm Shift within the Cognitive Sciences. In M. Lungarella, F. Iida, J. Bongard & R. Pfeifer (Eds.), *50 years of artificial intelligence: Essays dedicated to the 50th anniversary of artificial intelligence* (pp. 63–75). Berlin: Springer-Verlag. doi:10.1007/978-3-540-77296-5_7

Froese, T. (2010). From Cybernetics to Second-Order Cybernetics: A Comparative Analysis of Their Central Ideas. *Constructivist Foundations*, *5*(2), 75–85.

Froese, T., & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, *173*(3–4), 466–500. doi:10.1016/j.artint.2008.12.001

Füllsack, M. (2013). Author's Response: Constructivism as Possibility? *Constructivist Foundations*, *9*(1), 23–25.

Garrett, R. (1999). Epistemology. In B. A. Thyer (Ed.), *The Philosophical Legacy of Behaviorism* (pp. 69–88), Dordrecht: Springer.

Garson, J. (2015). Connectionism. *Stanford Encyclopedia of Philosophy*. Retrieved August 12, 2018, from https://plato.stanford.edu/entries/connectionism/.

Garvey, J. (2010). Hacker's challenge. Retrieved August 26, 2018, from http://www.philosophypress.co.uk/?p=1583

Geisler, W. S., & Diehl, R. L. (2003). A Bayesian approach to the evolution of perceptual and cognitive systems. *Cognitive Science*, *27*, 379–402. doi:10.1207/s15516709cog2703_3

Graham, P. J. (2014). The Function of Perception. In A. Fairweather (Ed.), *Virtue Scientia: Bridges between Virtue Epistemology and Philosophy of Science* (pp. 13–31). Berlin: Springer.

Green, C. D. (1998). Are Connectionist Models Theories of Cognition? *Psycoloquy*, *9*(4).

Hacker, P. M. S. (2001). An orrery of intentionality. *Language and Communication*, *21*(2), 119–141. doi:10.1016/s0271-5309(00)00016-1

Hamblin, S. (2013). On the practical usage of genetic algorithms in ecology and evolution. *Methods in Ecology and Evolution*, *4*(2), 184–194. doi:10.1111/2041-210x.12000

Hancock, S., Gareze, L., Findlay, J. M., & Andrews, T. J. (2012). Temporal patterns of saccadic eye movements predict individual variation in alternation rate during binocular rivalry. *i-Perception*, *3*(1) , 88–96. doi:10.1068/i0486

Hand, D. J. (2014). Wonderful Examples, but Let's not Close Our Eyes. *Statistical Science*, *29*(1), 98–100. doi:10.1214/13-sts446

Hart, W. E., & Belew, R. K. (1991). Optimizing an arbitrary function is hard for the genetic algorithms. In R. K. Belew & L. B. Booker (Eds.), *Proc. 4th Int. Conf. Genetic Algorithms, San Mateo* (pp. 190–195). CA: Morgan Kaufmann.

Harvey, I. (2000). Robotics: Philosophy of Mind Using a Screwdriver. In T. Gomi (Ed.), *Evolutionary Robotics: From Intelligent Robots to Artificial Life* (pp. 207–230). Ontario: AAI Books.

Haselager, W. F. G., de Groot, A. D., & van Rappard, J. F. H. (2003). Representationalism versus anti-representationalism: A debate for the sake of appearance. *Philosophical Psychology*, *16*(1), 5–23. doi:10.1080/0951508032000067761

Hawkins, S. L. (2011). William James, Gustav Fechner, and early psychophysics. *Frontiers in Physiology*, 2, 1–129. doi:10.3389/fphys.2011.00068

Heidelberger, M. (2004). *Gustav Theodor Fechner and His Psychophysical Worldview* (C. Klohr, Trans.). Pittsburgh, PA: University of Pittsburgh Press.

Hoffman, D. D. (2009). The interface theory of perception. In S. Dickinson, M. Tarr, A. Leonardis & B. Schiele (Eds.), *Object categorization: computer and human vision perspectives* (pp. 148–165). New York: Cambridge University Press.

Hoffman, D. D. (2011). The construction of visual reality. In J. Blom & I. Sommer (Eds.), *Hallucinations: theory and practice* (pp. 7–15). New York: Springer.

Hoffman, D. D. (2018). The interface theory of perception. In J. T. Wixted & J. Serences (Eds.), *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, Volume 2, Sensation, Perception, and Attention, 4th Edition* (pp. 731–754). Hoboken, NJ: Wiley.

Hoffman, D. D., & Singh, M. (2012). Computational evolutionary perception. *Perception*, *41*, 1073–1091. doi:10.1068/p7275

Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin & Review*, *22*(6), 1480–1506. doi:10.3758/s13423-014-0776-1

Hoffman, D. D., Singh, M., & Mark, J. (2013). Does evolution favor true perceptions? In B. E. Rogowitz, T. N. Pappas & H. de Ridder (Eds.), *Proceedings of the SPIE 8651, Human Vision and Electronic Imaging XVIII* (pp. 24–36). Springfield, VA: SPIE. doi:10.1117/12.2011609

Hutto, D. D. (2013). Radically enactive cognition in our grasp. In Z. Radman (Ed.), *The hand: an organ of the mind* (pp. 227–252). Cambridge, MA: MIT Press.

Hutto, D., & Myin, E. (2013). *Radicalizing enactivism: basic minds without content*. Cambridge: MIT Press.

Hutto, D.D. (2017). REC: Revolution Effected by Clarification. *Topoi*, *36*(3), 377–391. doi:10.1007/s11245-015-9358-8

Jonas, H. (1966). *The phenomenon of life: Towards a philosophical biology*. Evanston, IL: Northwestern University Press.

Kaznatcheev, A., Montrey, M., & Shultz, T. R. (2014). Evolving useful delusions: Subjectively rational selfishness leads to objectively irrational cooperation. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, *36*, 731–736.

Keeney, B. P. (1983). *Aesthetics of Change*. New York: Guilford Press.

Kirwan, C. B., Ashby, S. R., & Nash, M. I. (2014). Remembering and imagining differentially engage the hippocampus: a multivariate fMRI investigation. *Cognitive Neuroscience*, *5*(3–4), 177–185. doi:10.1080/17588928.2014.933203

Kjellman, K. A. (2003). *Constructive systems science – the only remaining alternative? A contribution to science and human epistemology* (Doctoral dissertation). Retrieved from http://www.diva-portal.org/smash/get/diva2:9397/FULLTEXT01.pdf

Klein, P. (2015). Skepticism. *Stanford Encyclopedia of Philosophy*. Retrieved July 22, 2018, from https://plato.stanford.edu/entries/skepticism/.

Knill, D., & Richards, W. A. (1996). *Perception as Bayesian inference*. New York: Cambridge University Press.

Kolenik, T. (2016a). Embodied Cognitive Robotics, the Question of Meaning and the Necessity of Non-Trivial Autonomy. In O. Markič, M. Gams, T. Strle & U. Kordeš (Eds.), *Cognitive science: proceedings of the 19th International Multiconference Information Society – IS 2016* (pp. 24–27). Ljubljana: Institut Jožef Stefan.

Kolenik, T. (2016b). Kaj imata skupnega Heidegger in Roomba: utelešena umetna inteligenca kot peskovnik kontinentalne filozofije. *Analiza*, *20*(3–4), 147–164.

König, B. (2001). Natural Selection. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 10392–10398). Oxford: Pergamon.

Kordeš, U. (2016). Going Beyond Theory: Constructivism and Empirical Phenomenology. *Constructivist Foundations*, *11*(2), 375–385.

Kuhn, T. S. (1959). The essential tension. Tradition and innovation in scientific research. In C. W. Taylor (Ed.), *The third (1959) University of Utah research conference on the identification of scientific talent* (pp. 162–174). Salt Lake City: University of Utah Press. (Reprinted in: Kuhn, T. S. (1977). The essential tension (pp. 225–239). Chicago: University of Chicago Press).

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding* (Cambridge Studies in Philosophy). Cambridge: Cambridge University Press.

Lehar, S. (2003). Gestalt isomorphism and the primacy of subjective conscious experience: a Gestalt Bubble model. *Behavioral and Brain Sciences*, *26*(4), 375–444. doi:10.1017/s0140525x03000098

Luhmann, N. (1988). *Erkenntnis als Konstruktion*. Bern: Benteli Verlag.

Luhmann, N. (1996). *Social Systems* (J. Bednarz, Jr. & D. Baecker, Trans.). Stanford, CA: Stanford University Press.

Mancilla, R. (2011). Introduction to Sociocybernetics (Part 1): Third Order Cybernetics and a Basic Framework for Society. *Journal of Sociocybernetics*, *42*(9), 35–56.

Mark, J. T. (2013). *Evolutionary pressures on perception: when does natural selection favor truth?* (Doctoral dissertation). Retrieved from http://antpac.lib.uci.edu/record=b5298832~S7

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Maturana, H. (1988). Reality: The search for objectivity, or the quest for the compelling argument? *The Irish Journal of Psychology*, *9*(1), 25–82. doi:10.1080/03033910.1988.10557705

Maturana, H. R. (1978). Biology of language: the epistemology of reality. In G. A. Miller & E. Lenneberg (Eds.), *Psychology and Biology of Language and Thought. Essays in Honor of Eric Lenneberg* (pp. 27–63). New York: Academic Press.

Maturana, H. R. (1987). Everything said is said by an observer. In W. I. Thompson (Ed.), *Gaia: A way of knowing* (pp. 65–82). New York: Lindisfarne Press.

Maturana, H. R., & Varela, F. J. (1988). *Tree of Knowledge: The Biological Roots of Human Understanding*. Colorado: Shambhala.

McMullin, B., & Varela, F. J. (1997). Rediscovering computational autopoiesis. In P. Husbands & I. Harvey (Eds.), *Fourth European Conference on Artificial Life* (pp. 38–47). Cambridge, MA: MIT Press.

Meyer, D., & Kieras, D. (1997). A computational theory of executive cognitive processes and human multiple-task performance: Part 1, basic mechanisms. *Psychological Review*, *104*(1), 3–65. doi:10.1037/0033-295x.104.1.3

Mingers, J. (1995). *Self-Producing Systems: Implications and Applications of Autopoiesis*. New York: Springer US.

Mirolli, M., & Parisi, D. (2011). Towards a Vygotskyan Cognitive Robotics: The Role of Language as a Cognitive Tool. *New Ideas in Psychology*, *29*(3), 298–311. doi:10.1016/j.newideapsych.2009.07.001

Mitchell, M. (1999). *An introduction to genetic algorithms*. Cambridge: The MIT Press.

Mitchell, M. (2011). *Complexity: A Guided Tour*. Oxford: Oxford University Press.

Möller, R. (1999). Perception through anticipation—a behavior-based approach to visual perception. In A. Riegler, M. Peschl & A. von Stein (Eds.), *Understanding Representation in the Cognitive Sciences* (pp. 169–176). New York: Plenum Academic / Kluwer Publishers.

Moravec, H. (1988). *Mind Children*. Cambridge, MA: Harvard University Press.

Moussaïd, M., Helbing, D., & Theraulaz, G. (2011). How simple rules determine pedestrian behavior and crowd disasters. *Proceeding of the National Academy of Sciences of the United States of America*, *108*, 6884–6888. doi:10.1073/pnas.1016507108

Noë, A. (2004). *Action in perception*. Cambridge: The MIT Press.

Ó Nualláin, S. (2002). *The Search For Mind: A new foundation for Cognitive Science*. Bristol: Intellect Books.

O'Regan, J. K., & Noë, A. (2001). A Sensorimotor Account of Vision and Visual Consciousness. *Behavioral and Brain Sciences*, *24*, 939–973. doi:10.1017/s0140525x01000115

Oyama, S. (1985). *The ontogeny of information: developmental systems and evolution*. Cambridge, Cambridge University Press.

Palmer, S. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Bloom-Lloyd (Eds.), *Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Lawrence Erlbaum Associates.

Palmer, S. E. (1999). *Vision science: photons to phenomenology*. Cambridge, MA: MIT Press.

Pecher, D., & Zwaan, R. A. (2006). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge: Cambridge University Press.

Peschl, M. F. (1991a). A simulation system for the investigation of cognitive processes in artificial cognitive systems-Radical connectionism and computational neuroepistemology. *IEEE International Joint Conference on Neural Networks*, *3*, 2210–2215. doi:10.1109/IJCNN.1991.170716

Peschl, M. F. (1991b). Interdisciplinarity and Self-Organization in Computational Neuroepistemology. In M. C. Jackson, G. J. Mansell, R. L. Flood, R. B. Blackham & S. V. E. Probert (Eds.), *Systems Thinking in Europe* (pp. 189–195). Boston, MA: Springer.

Peschl, M. F. (1999). The development of scientific theories and their embodiment in the representational activities of cognitive systems. Neural representation spaces, theory spaces and paradigmatic shifts. In P. V. Loocke (Ed.), *The Nature of Concepts* (pp. 184–214). London: Routledge & Kegan Paul.

Peschl, M. F., & Riegler, A. (1999). Does Representation Need Reality? In A. Riegler, M. Peschl & A. von Stein (Eds.), *Understanding Representation in the Cognitive Sciences* (pp. 9–17). New York: Kluwer Academic / Plenum Publishers.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.

Piaget, J., Gruber, H. E., & Vonèche, J. J. (1995). *The essential Piaget*. Northvale, NJ: J. Aronson.

Pinel, J. P. (2013). *Biopsychology*. Harlow, England: Pearson.

Putnam, H. (1968). Brains and behaviour. In R. J. Butler (Ed.), *Analytical philosophy, vol. II* (pp. 1–19). Oxford: Basil Blackwell.

Quine, W. V. O. (1969). Epistemology naturalized. In W. V. O. Quine (Ed.), *Ontological Relativity and Other Essays* (pp. 69–90). New York: Columbia University Press.

Riegler, A. (2001). Towards a Radical Constructivist Understanding of Science. *Foundations of Science*, *6*(1), 1–30. doi:10.1023/A:1011305022115

Riegler, A. (2008). Natural or internal selection? The case of canalization in complex evolutionary systems. *Artificial Life*, *14*(3), 345–362. doi:10.1162/artl.2008.14.3.14308

Riegler, A. (2012). Constructivism. In L. L'Abate (Ed.), *Paradigms in theory construction* (pp. 235–256). New York: Springer.

Riegler, A., Stewart, J., & Ziemke, T. (2013). Computational Approaches to Constructivism [Special issue]. *Constructivist Foundations*, *9*(1).

Rosen, R. (1991). *Life itself: A comprehensive enquiry into the nature, origin and fabrication of life*. New York: Columbia University Press.

Rumelhart, D. E., & McClelland, J. (1986). On learning the past tense of English verbs. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel Distributed Processing, Volume 2* (pp. 216–271). Cambridge, Mass: MIT Press.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, *76*, 677–694. doi:10.1016/j.neuron.2012.11.001

Searle, J. R. (1986). *Minds, Brains and Science*. Cambridge: Harvard University Press.

Shanahan, M. (2016). The Frame Problem. *Stanford Encyclopedia of Philosophy*. Retrieved July 23, 2018, from https://plato.stanford.edu/entries/frame-problem/.

Shultz, T. R., & Sirois, S. (2008). Computational models of developmental psychology. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 451–476). New York, NY: Cambridge University Press.

Sipser, Michael (1997). *Introduction to the Theory of Computation*. Boston, MA: PWS Publishing.

Skinner, B. E. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Steup, M. (2005). Epistemology. *Stanford Encyclopedia of Philosophy*. Retrieved July 23, 2018, from https://plato.stanford.edu/entries/epistemology/.

Stewart, J. (2000). From Autopoiesis to Semantic Closure. In J. L. R. Chandler & G. van der Vijver (Eds.), *Closure: Emergent Organizations and Their Dynamics* (pp. 155–162). Annals of the New York Academy of Sciences.

Sun, R. (2008a). Introduction to computational cognitive modeling. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 3–19). New York: Cambridge University Press.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science*, *25*(2), 203–244. doi:10.1207/s15516709cog2502_2

Sun, R. (2008b). Theoretical status of computational cognitive modeling. *Cognitive Systems Research*, 10, 124–140. doi:10.1016/j.cogsys.2008.07.002

Thagard, P. (2013). Cognitive science. Retrieved August 21, 2018, from https://www.britannica.com/science/cognitive-science

Thompson, E. (2007). *Mind in life: biology, phenomenology, and the sciences of mind*. Cambridge: Harvard University Press.

Thyssen, O. (2004). Luhmann and Epistemology. *Cybernetics & Human Knowing*, *11*(1), 7–22.

Tolk, A. (2013). *Ontology, Epistemology, and Teleology for Modeling and Simulation Philosophical Foundations for Intelligent M&S Applications*. Berlin: Springer-Verlag.

Varela, F. J., Maturana, H., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, *5*(4), 187–196. doi:10.1016/0303-2647(74)90031-8

Varela, F. J., Thompson, E., & Rosch, E. (2016). *The embodied mind: cognitive science and human experience*. Cambridge, MA: The MIT Press. (Original work published 1991)

Vernon, D. (2005). Cognitive Vision: The Development of a Discipline. *KI*, *19*, 43–46.

Villalobos, M., & Ward, D. (2015). Living Systems: Autonomy, Autopoiesis and Enaction. *Philosophy & Technology*, *28*(2), 225–239. doi:10.1007/s13347-014-0154-y

Von Foerster, H. (1962). *Principles of Self-Organization*. New York: Pergamon Press.

Von Foerster, H. (1995). Ethics and second-order cybernetics. *Stanford Humanities Review*, *4*(2), 308–319.

Von Foerster, H. (2003). *Understanding Understanding: Essays on Cybernetics and Cognition*. New York, NY: Springer-Verlag.

Von Glasersfeld, E. (1981). The Concepts of Adaptation and Viability in a Radical Constructivist Theory of Knowledge. In I. E. Sigel, D. M. Brodzinsky & R. M. Golinkoff (Eds.), *Piagetian theory and research* (pp. 87–95). Hillsdale, NJ: Erlbaum.

Von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. London: Falmer Press.

Von Uexküll, J. (1957). A Stroll Through the Worlds of Animals and Men: A Picture Book of Invisible Worlds. In C. H. Schiller (Ed.), *Instinctive Behavior: The Development of a Modern Concept* (pp. 5–80). New York: International Universities Press.

Von Uexküll, J. (1982). The Theory of Meaning. *Semiotica*, *42*(1), 25–82. doi:10.1515/semi.1982.42.1.25

Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The varieties of enactivism. *Topoi*, *36*(3), 365–375. doi:10.1007/s11245-017-9484-6

Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and Computing*, *4*(2), 65–85. doi:10.1007/BF00175354

Zárate-Miñano, R., Anghel, M., & Milano, F. (2013). Continuous wind speed models based on stochastic differential equations. *Applied Energy*, *104*, 42–44. doi:10.1016/j.apenergy.2012.10.064

12 Appendices

12.1 Appendix A: Alphabetical List of Used Abbreviations

Abbreviation	Full form	Abbreviation first appears in chapter:
AE	**a**utopoietic **e**nactivism	6.1
GA	**g**enetic **a**lgorithm	4
COKFLONE	**c**oncept **o**f the **k**nowledge **f**eedback **l**oop **o**f **n**atural **e**pistemology	1.2
REC	**r**adically **e**nactive **c**ognition	6.1
SE	**s**ensorimotor **e**nactivism	6.1
SEM	Hoffman et al.'s (2015) **s**ensory-**e**xclusive **m**odel	5.1
SMM	my **s**ensori**m**otor **m**odel	5.2/6

12.2 Appendix B: The MIT License

"Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE."

12.3 Appendix C: Links to the Models' GitHub Repositories

Hoffman et al.'s (2015) sensory-exclusive model (SEM): https://github.com/TineKolenik/hoffman_reproduction

My sensorimotor model (SMM): https://github.com/TineKolenik/hoffman_enactivist_upgrade

13 Extended Summary in Slovenian Language: Računalniško modeliranje vpliva naravnega izbora na veridičnost zaznavanja

13.1 Uvod

Vprašanje odnosa med materialnim svetom in umom je vprašanje brez odgovora. Kljub številnim teorijam marsikaterih ved in znanosti poenotenja ni. Filozofska epistemologija (Kvanvig, 2003) predstavlja nosilko znanja o tem vprašanju, s kognitivno revolucijo pa se je le-tega z znanstveno metodo lotila tudi kognitivna znanost (Ó Nualláin, 2002). Slednja se je skozi desetletja obstoja organizirala v svojevrstno disciplino, v kateri se vez med filozofijo, umetno inteligenco, nevroznanostjo, psihologijo, jezikoslovjem in ostalimi konstitutivnimi disciplinami kaže v nenehnem srečevanju raznolikega znanja o spoznavanju živih bitij in različnih predpostavk raziskovalk, ki spoznavanje živih bitij zasnavljajo, raziskujejo in interpretirajo. Mnoge raziskovalke menijo, da lahko računalniško modeliranje pomaga pri spopadanju s težkimi epistemološkimi vprašanji (Froese, 2007). Zaznavanje je bilo opredeljeno kot en glavnih virov spoznavanja o zunanjem svetu (Alston, 1999), evolucijski vidik pa morda nudi eno najzanimivejših obravnavanj tega fenomena. Računalniško modeliranje se je v preteklosti že uporabilo za raziskovanje evolucije zaznavanja (Hoffman, Prakash in Singh, 2015).

Magistrsko delo obravnava evolucijo zaznavanja, kjer je v ospredju vprašanje, kakšno zaznavanje prinaša največjo preživetveno vrednost organizmom. To raziskuje z računalniško metodo genetskih algoritmov (Mitchell, 1999). Slednje služi kot primer za širšo kritično analizo računalniškega modeliranja kot metode za raziskovanje odnosa med zunanjim svetom in umom ter kritično analizo vloge računalniškega modeliranja kot metode kognitivne znanosti kot naravne epistemologije. Delo želi pokazati, da je raba računalniških modelov precej manj preprosta, kot kaže siceršnja praksa.

13.2 Kognitivna znanost kot naravna epistemologija

Zdi se, da kognitivna znanost in filozofska epistemologija poskušata odgovarjati na izredno podobna, če ne celo ista vprašanja (Ó Nualláin, 2002), pri čemer je pri obeh v ospredju vprašanje odnosa med zunanjim svetom in umom (Kvanvig, 2003; Thagard, 2013). Kognitivna znanost in filozofska epistemologija obravnavata nekatere iste vire spoznavanja, kjer si zastavljata celo ista vprašanja. O zaznavanju obe med drugim zanima, kako preko njega spoznavamo, kakšna je vloga vzročnosti v njem ter kako zaznavanje in njegove objekte konceptualiziramo. Na ta vprašanja filozofska epistemologija odgovarja s filozofsko analizo (Berkeley, 1710/1982), kognitivna znanost pa med drugim z umetnim vidom znotraj discipline umetne inteligence (Marr, 1982). Obe metodi ponujata neke vrste odgovorov na ista vprašanja kognitivne znanosti in filozofske epistemologije. Tako se kaže, da med disciplinama obstaja globlja vez, ki se odseva tudi v zgodovinskem razmerju med obema, v katerem raziskovanje z znanstveno metodo nujno in pomembno temelji na filozofskih analizah iz bogate zapuščine epistemologije. Tako recimo Piagetovo delo močno temelji na delu Kanta (Fabricius, 1983), delo psihofizikov (Jamesa, Webra, Wundta, Fechnerja …) pa na delu Spinoze, Leibniza idr. To pojasnjuje tudi dejstvo, da so se pojavili glasovi, ki si prizadevajo prikazati povezavo med kognitivno znanostjo in filozofsko epistemologijo kot nekaj pomembnega, smiselnega in vrednega nadaljnjega razvoja (Bateson, 1979; Keeney, 1983; Luhmann, 1988; Quine, 1969). To povezavo Keeney (1983) poimenuje naravna epistemologija, pojavljajo pa se tudi izrazi empirična epistemologija (Luhmann, 1988)), eksperimentalna epistemologija (Ó Nualláin, 2002) in naturalizirana epistemologija (Quine, 1969). Mnoštvo definicij se lahko popreprosti na slednjo: Naravna epistemologija preučuje epistemološka vprašanja z naravnimi znanstvenimi metodami.

Za razumevanje naravne epistemologije je potrebno razumevanje njenega razvoja skozi kognitivno znanost. V kognitivni znanosti kot naravni epistemologiji se koncepta konvergentnega in divergentnega mišljenja (Kuhn, 1959) manifestirata v krožnem oplajanju védenja znotraj kognitivne znanosti, ki ga predstavljajo epistemološki uvidi o spoznavanju živih bitij, in védenja kognitivne znanosti, ki ga predstavlja epistemologija kognitivne znanosti kot discipline in tistih, ki znotraj nje raziskujejo. Védenje znotraj kognitivne znanosti predstavlja konvergentno mišljenje, ki se za Kuhna odvija takrat, ko znanstvenica ni inovatorka, temveč rešuje težave, ki so lahko zasnovane in rešene znotraj trenutne znanstvene paradigme, niso pa namenjene temeljnim odkritjem ter znanstvenim revolucijam. Divergentno mišljenje je tisto, ki pripelje do temeljnih odkritij in znanstvenih revolucij, to pa deluje na nivoju védenja kognitivne znanosti. Divergentno mišljenje je potrebno vedno, ko raziskovanje v obstoječi paradigmi trči ob zid, ko obstoječa orodja védenja o spoznavanju živih bitij ne ponujajo rešitve. Primer v zgodovini kognitivne znanosti predstavlja že poprej omenjeno raziskovanje vida znotraj umetne inteligence, v katerem je kognitivistična paradigma, ki postavlja um analogno informacijskoprocesnemu računalniku, zanemarjala vlogo senzomotorične zanke pri porajanju vidnega sveta, sprememba paradigme pa je bila potrebna za boljše razumevanje fenomena in posledično uspešnejšo implementacijo umetnega vida (Jug, Kolenik, Ofner in Farkaš, 2018). Zanka védenja znotraj kognitivne znanosti in védenja kognitivne znanosti odraža miselnost kibernetike in kibernetike drugega reda; slednja zagovarja, da je potrebno zavedanje, da žive sisteme raziskujejo prav živi sistemi sami, in da je potrebno žive sisteme, ki raziskujejo, tudi upoštevati. Ta uvid je pomemben del naravne epistemologije.

13.3 Evolucija zaznavanja

Namen zaznavanja naj bi bilo »tvorjenje notranje reprezentacije prostorske navidezne replike zunanjega sveta« (Lehar, 2003, str. 375). Takšno zaznavanje naj bi bilo posledica evolucijske usmeritve »razvoja organizmovega [...] sistema, da se ujema s strukturo sveta in kodirno shemo narave« (Knill in Richards, 1996, str. 6). To stališče beremo v univerzitetnih učbenikih (Friedenberg in Silverman, 2016; Palmer, 1999) in v temeljnih delih največjih znanstvenih avtoritet (Craik, 1943; Fodor, 1985; Marr, 1982). Sodobne bayesovske teorije, ki predstavljajo alternativo tradicionalnim, trdijo podobno, saj naj bi bile zaznavne ocene, ki natančno prikazujejo resnico, koristnejše od tistih, ki tega ne počnejo (Geisler in Diehl, 2003). To, da »imamo [...] *veridično zaznavanje*« (Palmer, 1999, str. 6; poudarek njegov), je v prevladujoči paradigmi kognitivne znanosti samoumevno in se predpostavlja (Bermúdez, 2014; Froese, 2007; Palmer, 1999). A prav Palmer (in večina njegovih somišljenic) hkrati trdi, da se je zaznavanje razvilo zato, da »služi organizmovemu preživetju in uspešnemu razmnoževanju« (Palmer, 1999, str. 5). Ali je namen kognicije torej izomorfna odslikava sveta ali služenje preživetju organizma? »Priznam, evolucije ne zanima veridičnost; narava kot taka ne izbira glede na resnico in natančnost. [...] A mislim, da zaznavna stanja služijo fitnesu tako, da natančno preslikavajo okolje [...],« pravi Graham (2014, str. 19). Ravno ta »mislim« odraža nevarno samoumevnost, skriva pa celo protislovno stališče. Zagovorniki veridičnega zaznavanja namreč hkrati trdijo, da »živi organizmi [zaznavajo] glede na njihove potrebe« (Falk, 2018, str. 99), saj morajo biti evolucijske »prilagoditve [...] razumljene kot sporazum med različnimi potrebami organizma« (König, 2001, str. 10395).

Zaznavanje v opisanem zastopa odnos med zunanjim svetom in umom. Za magistrsko delo je pomembno razumeti dve razlikovanji. Prvo razlikovanje je med *veridičnim* in *neveridičnim zaznavanjem*. *Veridično zaznavanje* označuje zaznavanje, ki natančno opiše zaznavano okolje. *Neveridično zaznavanje* je nasprotno veridičnemu. Neveridično zaznavanje je lahko *izomorfno* ali *neizomorfno*, veridično pa le *izomorfno*; to predstavlja drugo razlikovanje. *Izomorfizem* pomeni ohranjanje fizičnih struktur in relacij zunanjega sveta tudi v samem mentalnem sistemu (Haselager, de Groot in van Rappard, 2003).

Tovrstna težka vprašanja, kot je vprašanje, ali je zaznavanje veridično ali neveridično ter izomorfno ali neizomorfno, naj bi lahko raziskovali z računalniškim modeliranjem: »Umetni inteligenci je dán privilegij, da nam lahko pomaga razrešiti filozofske spore, ki pestijo zahodno filozofijo že desetletja, če ne stoletja« (Froese, 2007, str. 68). Tovrstnega raziskovanja se je lotil Hoffman (2018), ki s sodelavci (2015) poskuša potrditi oz. ovreči svojo teorijo zaznavnega vmesnika. Slednja trdi, da zaznavanje ne služi »resnici«, temveč evolucijskemu fitnesu, živa bitja pa imajo zaznavni dostop do sveta le preko vmesnika. Ta je analogen računalniškemu zaslonu: to, da je ikona določene barve in da jo lahko izbrišemo, ne predstavlja prave narave tega, kar se dogaja v ozadju (delovanje tranzistorjev, vezja, električnega toka itd.). Hoffman in sodelavci (2015) so med drugim teorijo preverili tudi z računalniškim modeliranjem z genetskimi algoritmi. Računalniški model po besedah avtorjev pokaže, da organizmi pod vplivom naravnega izbora razvijejo neveridično, neizomorfno zaznavanje.

Delo Hoffmana in sodelavcev je raziskovanju v pričujočem magistrskem delu služilo na dveh nivojih. Na prvem nivoju smo trditve, ki jih Hoffman in sodelavci postavijo na podlagi rezultatov računalniškega modela, preverili z modelom, ki deluje na drugih (epistemoloških) predpostavkah. Na drugem nivoju smo, s pomočjo naučenega iz dela na prvem nivoju, kritično analizirali vlogo računalniškega modeliranja znotraj kognitivne znanosti kot naravne epistemologije.

13.4 Raziskovalna vprašanja in cilji

Magistrsko delo je imelo dva raziskovalna cilja (RC). Prvi je bil splošen in vseprisoten v delu, drugi pa je bil potreben za končno realizacijo prvega.

RC1) Kritična analiza vrednosti in vloge računalniškega modeliranja (z genetskimi algoritmi) za razumevanje (epistemoloških) predpostavk v modelih kognicije ter vloge računalniškega modeliranja za preučevanje epistemoloških vprašanj naravne epistemologije. Izhodišče je bilo v izdelanih računalniških modelih, ki bosta služila razumevanju samega postopka iz prve roke.

RC2) Raziskovanje zaznavne veridičnosti z računalniškim modeliranjem z genetskimi algoritmi. Glavno vprašanje je bilo, ali ima za modeliran organizem večjo preživetveno vrednost izomorfno ali neizomorfno zaznavanje. Modeliranje je temeljilo na delih Mitchell (1999) in Hoffmana idr. (2015). RC2 je imelo tri raziskovalna vprašanja:

1. Ali reprodukcija modela Hoffmana idr. (2015) povrne enake rezultate kot model avtorjev?

   Hipoteza je bila, da bodo rezultati reproduciranega modela enaki kot rezultati modela avtorjev.

2. Katere predpostavke vsebuje model Hoffmana idr. (2015)?

   Hipoteza je bila, da so predpostavke modela Hoffmana idr. (2015) analogne nekaterim predpostavkam kognitivizma.

3. Kako je lahko model Hoffmana idr. (2015) modificiran tako, da se nekatere kognitivistične predpostavke, ki jih vsebuje, zamenjajo z enaktivističnimi ter kako to vpliva na rezultate?

   Hipoteza je bila, da bodo rezultati enaki, vendar je bila sama modifikacija z enaktivističnimi predpostavkami težavna in nujna širše razprave.

13.5 Metode in postopki

Za odgovor na RC1 je bila potrebna obširna kritična analiza, ki je temeljila na karakterizaciji naravne epistemologije iz predhodnih poglavij ter na računalniškem modeliranju modela Hoffmana idr. (2015) in lastnega modela.

Za odgovor na RC2 je bila uporabljena metoda računalniškega modeliranja s tehniko genetskih algoritmov (Mitchell, 1999), ki so bili napisani v programskem jeziku Python. Genetski algoritem je optimizacijska hevristika, ki pri problemih, kjer ne poznamo najboljše rešitve ali pa bi bilo zanjo potrebna prevelika računska moč, vrne vsaj zadovoljive rešitve. Pomembnost genetskih algoritmov v raziskovanju naravnih fenomenov temelji na inherentnem vsebovanju evolucijskih mehanizmov naravnega izbora (izbira organizmov z najvišjim fitnesom glede na podano funkcijo), dedovanja (nov organizem je ustvarjen s presekom DNK dveh prednikov) in mutacije (določena je majhna verjetnost, da se gen spremeni v naključno alternativo iz bazena možnih genov).

13.6 Teorija zaznavnega vmesnika in senzorični model

Hoffman idr. (2015) Hoffmanovo teorijo zaznavnega vmesnika preverjajo na različne načine: z genetskimi algoritmi, z evolucijskimi igrami, z bayesiansko statistiko itd. Računalniški model z genetskimi algoritmi, ki je bil reproduciran, temelji na modelu Mitchell (2009), vsebuje pa robotka Robbyja, ki se premika po prostoru in poskuša zbrati karseda veliko število pločevink pijače. Robby lahko na mreži, ki jo tvori 10 × 10 kvadratov (oz. mest) in jo obkrožuje zid, vsak krog naredi eno od sedmih potez: premik severno, premik južno, premik vzhodno, premik zahodno, naključen premik, brez premika in poskus pobiranja pločevinke. Robby izgubi eno točko, če poskusi pobrati pločevinko, vendar te ni na mestu, kjer se Robby nahaja; izgubi pet točk, če se premakne z mesta proti zidu, s čimer se vanj zaleti; dobi 10 točk, če poskusi pobrati pločevinko in mu to tudi uspe, saj se ta nahaja na mestu, kjer Robby stoji. Robby »vidi« v tem svetu naslednje: mesto, na katerem stoji, mesto severno, mesto južno, mesto vzhodno in mesto zahodno, torej pet mest (oz. kvadratov) v svoji okolici. Na takšnem mestu je lahko ali pločevinka ali zid ali pa je mesto prazno. Robby tako lahko potencialno »vidi« 243 ali 3^5

situacij (tri možna stanja vsakega mesta, ki ga Robby »vidi«, mest pa je pet). Robby vsebuje verigo DNK, ki narekuje njegove poteze. Vsebuje informacije, kaj naj Robby v določeni situaciji stori. Vsaka izmed potencialnih 243 situacij je povezana z eno izmed sedmih potez, ki jih Robby lahko stori, kar pomeni, da ima njegov DNK 243 genov. Ta povezava med situacijami in potezami določa njegovo vedênje v svetu ter determinira končno število točk, ki jih nabira s pločevinkami. Genetski algoritem na začetku ustvari množico Robbyjev z različnimi verigami DNK. Vsak Robby ima svoj odziv z eno od sedmih potez na eno izmed 243 situacij, kjer so poteze in situacije naključno povezane. S tem dobimo Robbyje, ki so različno sposobni nabiranja pločevink. Izbranci z najboljšimi strategijami so nato dani v genski bazen, iz katerega se tvori novo generacijo Robbyjev, kjer ima vsak novi Robby DNK, ki je plod dveh polovic DNK dveh Robbyjev iz genskega bazena. Zgodi se lahko tudi genska mutacija, kar pomeni, da se lahko neki situaciji pripadajoča poteza spremeni v neko drugo potezo. V tem modelu so na začetku Robbyji nesposobni, zaletavajo se v zid in pobirajo pločevinke na mestih, kjer jih ni. A skozi generacije postajajo vse spretnejši v svoji strategiji, dokler se ne zaletavajo več ter pobirajo pločevinke tam, kjer tudi so.

Hoffman idr. (2015) spremenijo model tako, da Robbyju poleg DNK, ki predstavlja strategijo nabiranja, dodajo še eno verigo DNK, ki predstavlja strategijo zaznavanja. Spremenijo tudi svet okoli Robbyja, saj imajo zdaj mesta od nič do 10 pločevink. Strategija zaznavanja pripiše vsaki količini pločevink (0–10) eno barvo, rdečo ali zeleno. Robby nabira točke po naslednjem ključu, ki je del t. i. kriterijske funkcije (ang. *fitness function*, odseva evolucijski fitnes modeliranega organizma): 0 pločevink pomeni 0 točk, 1 pločevinka pomeni 1 točko, 2 pločevinki pomenita 3 točke, 3 pločevinke pomenijo 6 točk, 4 pločevinke pomenijo 9 točk, 5 pločevink pomeni 10 točk, 6 pločevink pomeni 9 točk, 7 pločevink pomeni 6 točk, 8 pločevink pomeni 3 točke, 9 pločevink pomeni 1 točko in 10 pločevink pomeni 0 točk. Takšen ključ pomeni torej nemonotono, gaussovsko porazdelitev točk glede na količino, avtorji pa ga upravičijo z razlago, da malo pločevink pijače pomeni dehidracijo, večje število pločevink pa utopitev, hkrati pa naj bi bili procesi homeostaze nemonotoni. Hoffmana idr. (2015) zanima, kakšna strategija zaznavanja se bo skozi generacije v modelu razvila. Če bi se razvila izomorfna strategija, bi imela DNK strategije zaznavanja z eno barvo označena mesta z manjšo količino pločevink, z drugo barvo pa mesta z večjo količino pločevink, s čimer bi zaznavanje Robbyju prenašalo informacije o zunanjem svetu (o količini pločevink zunaj Robbyja). Če bi se razvila neizomorfna strategija, bi imela DNK strategije zaznavanja z eno barvo označeno mesto količine pločevink, ki prinaša večje število točk za Robbyja, z drugo barvo pa mesto količine pločevink, ki prinaša manjše število točk za Robbyja, s čimer bi zaznavanje Robbyju prenašalo informacije o njegovih notranjih potrebah za preživetje (o točkah, ki jih dobi glede na določeno količino pločevink). Robby skozi generacije – model teče 500 generacij, v vsaki generaciji je 100 krogov, ko Robby potuje po mreži in nabira pločevinke – tako ne razvija le DNK strategije nabiranja, temveč tudi DNK strategije zaznavanja. Ko postane dober nabiralec, se izkaže, da razvije neizomorfno DNK strategije zaznavanja. Ta je ena izmed dveh možnosti: Robby »vidi« mesta s številom pločevink, ki mu prinaša malo točk, rdeče, mesta s številom pločevink, ki mu prinaša veliko točk, pa zeleno; ali pa »vidi« mesta s številom pločevink, ki mu prinaša malo točk, zeleno, mesta s številom pločevink, ki mu prinaša veliko točk, pa rdeče.

Model, ki ga poimenujem senzorični model, uspem reproducirati in dobim enak rezultat: Robbyji razvijejo neizomorfno zaznavanje, konkretno se v enem izmed primerov razvije naslednji DNK strategije zaznavanja: 0, 1, 9 in 10 pločevink je povezanih z eno barvo, vmesne količine pločevink pa z drugo barvo. Tako je potrjena hipoteza prvega raziskovalnega vprašanja RC2.

Analiza modela razkrije, da vsebuje nekatere kognitivistične predpostavke. To neposredno izrazijo tudi avtorji z dvema trditvima:

1. »Teorija zaznavanja preko vmesnika [ne zavrača informacijskega procesiranja]. Dokazov za informacijsko procesiranje je več kot dovolj.« (Hoffman et al., 2015, str. 1501)
2. »Trdimo, da motorika ni potrebna za zaznavo.« (prav tam, str. 1497)

Druga trditev se še posebej močno izraža v modelu avtorjev, saj njihov Robby zaznava nediskriminatorno, akcija pa nima neposrednega vpliva na to, kako zaznava in kako je to zaznavanje omejeno ter pogojeno z akcijo. Tako je potrjena hipoteza drugega raziskovalnega vprašanja RC2.

13.7 Enaktivizem in senzomotorični model

Analiza enaktivizma in možnosti njegovega modeliranja pokaže, da je najbolj smiseln element enaktivizma za modeliranje, tako v splošnem kot v primeru modifikacije modela Hoffmana idr. (2015), senzomotorična zanka. Ima dokumentirano zgodovino rabe v računalniških modelih (Froese in Ziemke, 2009), hkrati pa predstavlja koncept, proti kateremu Hoffman idr. (2015) eksplicitno nastopijo – po njihovem motorika ni potrebna za vidno zaznavo. Bistvo senzomotorične zanke je, da ima agent preko motorike kontrolo nad tem, kaj zazna, s čimer se gibanje in zaznavanje sodoločata do te mere, da ni jasno, kje je začetek in kje konec medsebojnega vplivanja.

Senzorični model je modificiran s senzomotorično zanko tako, da Robby ne »vidi« petih mest, temveč le dve – svoje in še eno. To, katero mesto še »vidi«, je odvisno od tega, kam je Robby po opravljeni potezi obrnjen, kar vpliva na naslednji krog, ko to, kam je obrnjen, določa to, kar »vidi«. Robby se torej poleg poteze odloči še, kam bo po njej »pogledal«: severno, južno, vzhodno ali zahodno. Mesto proti tej smeri neba mu je vidno v naslednjem krogu. Situacija para mesta, na katerem stoji, in mesta, proti kateremu je obrnjen, določi njegovo potezo. S tem je Robbyjevo zaznavanje omejeno in hkrati neizbežno pogojeno z njegovo aktivnostjo. Pri tem se tvori prava senzomotorična zanka, kjer sami krogi Robbyjevega delovanja niso več jasno zamejeni, s čimer je nemogoče pokazati, kje se začne akcija in konča zaznavanje ter obratno.

Rezultati potrdijo hipotezo tretjega raziskovalnega vprašanja RC2. Robbyji v modelu, ki je poimenovan senzomotorični model, razvijejo neizomorfno zaznavanje kljub drugačnim (epistemološkim) predpostavkam. Ponovno se DNK strategije zaznavanja razvije v eno od dveh možnosti: 0, 1, 9 in 10 pločevink je povezanih z rdečo, vmesne količine pločevink pa z zeleno oz. obratno, kar se tiče barv.

13.8 Primerjava modelov in dodatni eksperimenti

Modela kljub nekaterim različnim predpostavkam pokažeta, da je za Robbyje preživetveno najboljše neizomorfno zaznavanje. Robbyji dosežejo zelo podoben lokalni maksimum (~120 točk), največja razlika pa se skriva v potrebnih generacijah, da Robbyji v posameznem modelu to številko dosežejo. Robbyji v senzoričnem modelu potrebujejo približno 250 generacij za lokalni maksimum, Robbyji v senzomotoričnem modelu pa le približno 25 generacij. Ob analizi se zdi, da je povzročitelj tega fenomena prav senzomotorična zanka, ki deluje kot nekakšen optimizator. Senzomotorična zanka poviša varianco med Robbyji, kjer imajo najboljši Robbyji že v prvi generaciji višje število točk, ti pa tako povzročijo hitrejšo evolucijo proti lokalnemu maksimumu, saj je večja verjetnost, da so izbrani in da prenesejo svoje gene v naslednjo generacijo. Senzomotorična zanka pomeni, da so Robbyjeve odločitve pomembnejše – njegovi akcija in zaznavanje pomembneje vplivata drug na drugega, kar se izraža v tesnejši povezavi in večjem vplivu na nabiranje točk. Vpliv senzomotorične zanke ni bil pričakovan, kar govori v prid njeni legitimnosti.

V dodatnih eksperimentih je bilo v Robbyje v modelih prednastavljeno ali izomorfno ali neizomorfno zaznavanje. Njihov DNK strategije zaznavanja je bil fiksen, kar pomeni, da se ni razvijal, ni bilo ne rekombinacij ne mutacij, temveč so imeli Robbyji v vseh 500 generacijah enega od dveh možnih DNK strategije zaznavanja. Tovrstno eksperimentiranje je služilo dodatni potrditvi, da neizomorfno zaznavanje Robbyjem prinese več točk. Rezultati niso bili presenetljivi. Robbyji v senzoričnem modelu s prednastavljenim neizomorfnim zaznavanjem so dosegli približno 100 točk, s prednastavljenim izomorfnim zaznavanjem pa približno 25 točk. Robbyji v senzomotoričnem modelu s prednastavljenim neizomorfnim zaznavanjem so dosegli približno 100 točk, s prednastavljenim izomorfnim zaznavanjem pa približno 90 točk. Razlika v modelih ob izomorfnem zaznavanju je bila velika, analiza pa kaže na senzomotorično zanko kot povzročiteljico te razlike. Senzomotorična zanka je ponovno povzročila hitrejši razvoj proti lokalnemu maksimumu točk pri senzomotoričnem modelu.

13.9 Računalniško modeliranje kot metoda naravne epistemologije

Oznaka »računalniško« (oz. »računsko«, ang. *computational*) nosi s seboj obsežno in predvsem nerazčiščeno »prtljago«, saj se beseda ne uporablja vedno na isti način, kar povzroča dvoumnost in konceptualno zmedo. Riegler, Stewart in Ziemke (2013) opozorijo na to problematiko, ki se jim zdi ključna pri razumevanju računalniškega modeliranja, in oznako opredelijo v dveh smislih. V prvem smislu »računalniško« označuje modelirane procese, ki so tudi v svoji naravi konceptualizirani kot taki. K temu spada računalniška teorija uma, ki trdi, da um deluje kot računalnik. V drugem smislu »računalniško« označuje modele, ki so uporabljeni kot orodja za razmišljanje, predstavljanje, razumevanje in osvetljevanje raziskovanih fenomenov, ne pomenijo pa, da je modelirano samo po sebi računalniško, da je njihov jezik delovanja matematika. Tako lahko npr. vedênje vetra zelo natančno računalniško modeliramo, pri čemer funkcionalno razumemo, kaj se dogaja, hkrati pa napovedujemo vedênje vetra. A to ne pomeni, da veter sam po sebi izvaja račune, da in ko piha v gozdu. Težava sicer nastopi, ker ni jasno, kaj so pravila, ko modeliramo v prvem in drugem smislu hkrati, če sledimo teoriji računalniške teorije uma in modeliramo agenta v okolju, ki samo na sebi ne izvaja računov za delovanje. A večja težava, verjetno ena največjih težav pri modeliranju živih bitij, sicer tudi povezana z »računalniškim«, nastopi že pri modeliranju samega agenta, ne glede na to, v kakšnem smislu »računalniškega« je ta modeliran. To težavo Riegler (2001) poimenuje Pacmanov sindrom, ki trdi, da je neuporabno modelirati agente, kjer jim stvariteljica že v naprej definira pravila delovanja ter življenjske koncepte, kot sta npr. »hrana« in »sovražnik«. Tovrstni koncepti so smiselni za stvariteljico, ne za agenta, saj niso lastne njemu samemu. Védenje modeliranih agentov (npr. kaj je »hrana«) izhaja iz stvariteljice, programerke, raziskovalke, ne agenta samega. To je ključna težava pri modeliranju agentov, ki naredi računalniško modeliranje močno odvisno in pogojeno od same raziskovalke (Kjellman, 2013). Peschl in Riegler (1999) zato zagovarjata omejeno, a vseeno izredno moč računalniškega modeliranja. Njegovo vrednost ne vidita toliko v tehničnih podrobnostih, temveč v konceptualni vrednosti, ki pomaga pri razmišljanju o problemih in fenomenih, še posebej takrat, ko lahko računalniški model razkriva pomembnost predpostavk in strategij v raziskovanju ter vloge epistemološkega okvirja za razumevanje fenomena.

Uvid, da je fenomen lahko računalniško modeliran, kljub temu da sam po naravi ni računski, terja tudi premislek o legitimnosti modeliranja paradigm kognitivne znanosti, ki nastopajo proti računalniškim teorijam uma. Ena izmed takšnih paradigm je enaktivizem, na katerem delno temelji tudi senzomotorični model. Izkaže se, da je računalniško modeliranje enaktivističnih vidikov popolnoma smotrno, ko upoštevamo »računalniško« v drugem smislu po klasifikaciji Rieglerja idr. (2013). Še več, enaktivistične ideje so bile v preteklosti že mnogokrat modelirane (Froese in Ziemke, 2009), uvidi pa so vplivali na sam razvoj paradigme in obratno, enaktivistične ideje so močno vplivale na sam razvoj metode računalniškega modeliranja (Brooks, 1991). Ta medsebojni vpliv med enaktivizmom in računalniškim modeliranjem spominja na medsebojni vpliv med védenjem znotraj kognitivne znanosti in védenjem kognitivne znanosti. Tu tudi tiči možnost presojanja vrednosti računalniškega modeliranja kot metode za naravno epistemologijo. Računalniško modeliranje mora biti postavljeno v zanko védenj. Ko se to zgodi, je jasno, da je imelo računalniško modeliranje izredno pomemben vpliv na epistemološke uvide v kognitivni znanosti in na razvoj epistemologije kognitivne znanosti. Na epistemološke uvide v kognitivni znanosti je računalniško modeliranje vplivalo na vseh štirih nivojih analize (Sun, 2008a), ki so povezani s kognitivno znanostjo – znotraj sociološke analize (nivo interakcij med agenti), psihološke analize (nivo posameznega agenta), komponentne analize (nivo interakcij znotraj agentov) in fiziološke analize (nivo substratov agenta). Na razvoj epistemologije kognitivne znanosti je računalniško modeliranje vplivalo ob vsakem klasificiranem vzpostavljanju nove paradigme (Froese, 2010) – kognitivistične paradigme (um kot računalnik), konekcionistične paradigme (raba nevronskih mrež kot uvid v možgansko distribuiranost kognicije), paradigme utelešene in umeščene kognicije (raba robotov kot uvid v pomembnost senzomotoričnega sklapljanja z okoljem) ter enaktivistične paradigme (kjer so bili sámo računalniško modeliranje ter njegove težave ključni za vzpostavitev »protiračunske« ideje kognicije). Z razumevanjem umeščenosti računalniškega modeliranja v razvoj kognitivne znanosti kot naravne epistemologije ter globljo osvetlitvijo metode kot take sta senzorični in senzomotorični model lahko bila šele kritično analizirana.

13.10 Kritična analiza senzoričnega in senzomotoričnega modela

Hoffman idr. (2015) o rezultatih senzoričnega modela zapišejo, da jasno kažejo, da naravni izbor spodbuja razvoj neizomorfnega zaznavanja, ki je pogojen z evolucijskim fitnesom živih bitij, in zavrže izomorfno zaznavanje, ki odseva strukture zunanjega sveta. Takšna trditev ima dve težavi. Ker Hoffman idr. trdijo, da živa bitja zaznavajo neizomorfno in da to zaznavanje odseva njihove preživetvene potrebe, to pomeni, da tudi Hoffman idr. zaznavajo neizomorfno in to zaznavanje odseva njihove preživetvene potrebe. Kibernetika drugega reda opozarja, kako pomembno je to spoznanje, in če se le-to umisli do skrajnosti, njihova raziskava pomeni, da rezultati modela in hkrati teorija zaznavnega vmesnika ne zrcali »resnice« o svetu, temveč zrcali preživetvene potrebe njenih raziskovalcev. S tem bi se strinjal tudi Peschl (1999), ki predlaga, da znanstvene teorije ne služijo objektivnemu opisu sveta, temveč zmožnosti človeškega preživetja v okolju. Če za trenutek odmislimo uporabo rezultatov o živih bitjih na raziskovalcih samih, je potrebno razjasniti vlogo modelov kot potencialnih potrdil o tem, kako naj bi živa bitja zaznavala. Za takšno razjasnitev je potrebna analiza modelov na petih področjih: razlagalnost, napovednost, kompleksnost, abstraktnost in viabilnost genetskih algoritmov kot metode naravne epistemologije.

Vloga modelov je bila predstavljena kot razlagalna za razliko od napovedne. Modeli ponujajo funkcionalno razlago za možnost neizomorfne zaznave in kako bi se ta manifestirala, ki je v disciplinah, ki raziskujejo zaznavanje, prevladujoče povsem nemisljiva. Modela nista mišljena kot napovedna. Modela sta bila označena za preprosta, saj implementacijsko posedujeta majhno število vrstic programske kode, majhno število parametrov ter zelo omejeno sposobnost agentov. Ker sta pokazala iste rezultate, je verjetnost velika, da je to posledica preprostosti, saj bi šele večja kompleksnost ter odmik od pomembnosti kriterijske funkcije, ki določa, kaj je za agente dobro in kaj slabo, lahko povzročila drugačne rezultate s spreminjanjem predpostavk. Vseeno obstaja možnost, da rezultatska konvergenca v neizomorfnem zaznavanju kaže na determinirajoč vpliv notranjih preživetvenih potreb na zaznavanje za razliko od struktur zunanjega sveta kot takih. Modela sta bila označena za abstraktna, saj naslavljata fenomen v karseda splošnem smislu, brez konkretizacije znotraj nivojev analize (npr. katera živa bitja).

Viabilnost genetskih algoritmov kot metode naravne epistemologije temelji na analizi dveh točk. Prva zadeva arbitrarno naravo dela parametrov in mehanizmov, ki so namenjeni optimizaciji algoritma, saj ne temeljijo na bioloških atributih (bodisi zaradi neraziskanosti v biologiji bodisi zaradi služenja izključno tehničnim vidikom algoritma). Druga zadeva parametre in funkcije, ki naj bi odslikavali zunanji svet in »resnico« o živih bitjih, kar bržkone predstavlja večjo težavo kot jo morda prva točka. Arbitrarna narava nekaterih parametrov in mehanizmov, kot so verjetnost za mutacijo, verjetnost za prenos genov in točka rekombinacije v DNK, dobro služi takrat, ko metoda genetskih algoritmov ni uporabljena za raziskovanje naravnih fenomenov, saj arbitrarnost omogoča manevrski prostor za optimizacijo. Ko je metoda uporabljena za raziskovanje naravnih fenomenov, je vloga arbitrarnosti nejasna in zlahka izrabljena za doseg želenih rezultatov, saj izraža najmanj določeno pristranskost v interpretaciji in izbiranju bioloških atributov, na katerih potem temelji zasnova modela (Hart & Belew, 1991). A globlja težava se skriva v delih modela, ki so nastavljeni z mislijo, da se modelirajo lastnosti objektivnega sveta. Tu je ponovno aktualen Pacmanov sindrom, ki ga v primeru genetskega algoritma lahko identificiramo v kriterijski funkciji. Ta določa, kaj je za agenta dobro ali slabo, in posledično se v njej izraža vsiljenost delovanja in ciljev (ali vsaj pravil za vznik le-teh) raziskovalke za razliko od samodoločanja agenta samega. To je še posebej razvidno v senzoričnem in senzomotoričnem modelu, ko kriterijska funkcija ostaja vseskozi enaka, statična, hkrati pa najpomembneje določa evolucijo zaznavanja do te mere, da nanj ne vplivajo ostale predpostavke v modelih. Korak v smer »osvobajanja« agentov od svojih stvariteljic bi lahko bil v dinamični kriterijski funkciji, saj se interni cilji živih bitij vedno razvijajo, znova vzpostavljajo in s tem nenehno spreminjajo (Oyama, 1985). Tako bi se kriterijska funkcija nepričakovano spreminjala v kompleksnem sistemu začetnih pravil, s čimer bi začetna pogojenost stvariteljev izgubila svojo določno moč, stvaritelji sami pa ne bi bili več vsevedni o evolucijski trajekciji agenta.

Kljub številnim težavam se zdi, da je računalniško modeliranje edino znanstveno orodje, kjer raziskovalka pozna svet in agenta v njuni popolnosti. Ima celovit dostop do poznavanja enega in drugega, kar je za epistemološko raziskovanje sicer nepredstavljivo. Tukaj to, da model popolnoma reflektira raziskovalko, saj ta diktira model kot tak, ni več šibkost, temveč vrlina. Senzorični in senzomotorični model morata biti zato razumljena v tej luči. Ključni sta dve vprašanji: Ali se lahko iz modelov naučimo kaj o kogniciji? Če se lahko, je pristop modeliranja modelov z različnimi predpostavkami in njihova primerjava, ko se raziskuje nek fenomen, koristen? Odgovor na drugo vprašanje je lažji. Primerjava senzoričnega in senzomotoričnega modela se je že izkazala za zanimivo, saj kaže na pomemben vpliv senzomotorične zanke na razvoj Robbyjev. Zanka je bila dodana brez napovedi o vplivu, le-ta pa je bil tako povsem nepričakovan. Ujemanje z enaktivistično idejo o pomembnosti senzomotorične zanke je torej zelo pomenljivo. Na prvo vprašanje je odgovor pritrdilen, če modela dovolj dobro predstavljata bodisi prvi bodisi drugi smisel »računalniškega«, kot ga klasificirajo Riegler idr. (2013). Senzomotorična zanka se je že izkazala za dobro orodje za optimizacijo, četudi morda modela ne predstavljata dobro tega, kar modelirata. A o odgovoru na to vprašanje je bilo lažje misliti, ko smo oba modela, podobno kot poprej, postavili v zanko védenja znotraj kognitivne znanosti in védenja kognitivne znanosti ter premislili, kakšen vpliv bi modela lahko imela.

Modela epistemološkim uvidom v kognitivni znanosti ponujata najmanj funkcionalno razumevanje fenomena neizomorfnega zaznavanja. Kažeta, kako bi se zaznavanje razvijalo, če bi bila kriterijska funkcija takšna, kot se jo predpostavlja. Seveda ima sama kriterijska funkcija veliko pomanjkljivosti (npr. statičnost), vendar je uvid vseeno legitimen in uporaben. Senzomotorični model prav tako ponuja uvid v pomembnost senzomotorične zanke, kar kaže na večjo legitimnost enaktivističnih idej – to pa že meji na vpliv na epistemologijo kognitivne znanosti.

Modela bi lahko bila del več epistemoloških premikov kognitivne znanosti, ki zadevajo enaktivizem (senzomotorična zanka), avtonomijo (kot se jo razume v Pacmanovem sindromu), radikalni konstruktivizem (kognicija kot odsev notranjih potreb, ne zunanje »resničnosti«), kibernetiko drugega reda (pomembnost raziskovalk v raziskovanju) in pomen znanosti (služenje preživetju, ne objektivnemu opisu zunanjega sveta). A težko je govoriti o premikih, saj so se vse naštete domene že dogodile in so že del epistemologije kognitivne znanosti tako ali drugače. Zdi se, da je torej nemogoče napovedati premik in smer razvoja, ki bi bil popolnoma lasten modeloma, verjetno pa je tudi, da modela epistemoloških premikov ne bosta povzročila.

13.11 Razprava in zaključek

Magistrsko delo odpira veliko raziskovalnega prostora za prihodnje delo. Velik napredek bi bil storjen, če bi bila kriterijska funkcija spremenjena iz statične v dinamično, s čimer bi se omilil Pacmanov sindrom. Zanimiva bi bila tranzicija z izključno programskega modela na robotski model, kar bi naredilo definiranje kriterijske funkcije toliko težje in zato tudi bolj pomenljivo, ko bi se preučevalo delovanje modelov. Modela bi bilo vredno narediti bolj kompleksna in potrebno bi bilo premisliti, kako razširiti predpostavke in njihovo integracijo v Robbyje tako, da bi lahko imele vpliv na razvito zaznavanje (četudi bi se izkazalo, da se vseeno razvije neizomorfno zaznavanje). Teoretično je razvoj koncepta kognitivne znanosti kot naravne epistemologije (sploh zanke védenj) potrebno za nadaljnje razumevanje metod, ki ji pripadajo.

Magistrsko delo nima enega samega bistva, kar ga naredi razpršenega in bolj zapletenega za koherentno strukturiranje in predstavitev. Zaradi občasne nebrzdane ambicije in nezmožosti zamejevanja pričujoča množičnost idej, konceptov, področij, disciplin, pristopov in nivojev kljub tkanju rdeče niti in vseskoznemu povezovanju včasih ne deluje. Na koncu je vseeno jasno, da računalniško modeliranje nikoli ni le računalniško modeliranje. Metoda ni enoznačna in enostavna, kot se jo prepogosto obravnava in uporablja. Zdi se, da je predstavljenih več težav (npr. Pacmanov sindrom), kjer je zid navidezno nepremostljiv in je zato, pomenljivo, bržkone potreben premik v epistemologiji kognitivne znanosti za možnost napredka in razrešitve. A do takrat so nebrzdani poskusi rušenja nepremostljivih zidov, kot ga predstavlja pričujoče delo, nujni.