The General Theory of Consciousness

The Abstract Definition of the Processes Required for the Emergence of Consciousness

(Master's Thesis)

Emile Michel Hobo
University of Twente
Department of EWI
Research Group of Human Media Interaction
P.O. Box 217
7500 AE Enschede
The Netherlands
E-mail: hobo@hoboart.nl

Supervisors:
Anton Nijholt
Rieks op den Akker
Dirk Heylen
Mannes Poel

August 13^{th}, 2004

This thesis is dedicated to all true universal scientists.

Any scientist should be a rebel in order to properly break with formerly accepted but false ideas and come up with new ones. (And there's no harm in showing that you are.)

Foreword

Seen from my background in (what some people would call) civil life I'm very much interested in doing research on consciousness. Better understanding the nature of ourselves as well as the world around us helps form a stronger basis for interacting with that world. It helps to better understand motives and the principle of interaction itself. I'm very glad and grateful that I was permitted to do my thesis on the subject of consciousness. I'm quite sure that it will help understand the basic methods behind solving problems better. Hopefully the proposed framework also leads to solutions within human as well as media related research areas.

There's quite a list of people whom I have to thank. Of course I have to thank all the researchers who have done any valuable work in the past. I cannot name all these people explicitly here, since there are way too many. There's also a list of other people who I need to thank for reading my thesis and providing me with the necessary comments. These people are (in alphabetical order by last name) Rieks op den Akker, Dirk Heylen, Riccardo Manzotti, Anton Nijholt, Mannes Poel and Auke Pols.

There are also some other people whom I'd like to thank. These people have influenced my work by either making it possible or setting me forth on my ideas in some other way. The latter usually happened because they already implemented some of my other ideas and could even indicate what they had done on top of that. This most fortunately gave me an idea of where to start. These people are my parents, my sister and my brother, David Andel, Peter Eggenberger Hotz, Gabriel Gómez, Martin F. Krafft, Chandana Paul, Rolf Pfeifer, Harri Valpola and everybody else who's included in my literature list.[1]

[1] If you don't know why you're here, you must have said something about consciousness somewhere which implicitly helped form my idea.

Contents

List of Figures

List of Acronyms

AI	artificial intelligence
BES	basic emotional states
BIRU	Basic Intentional-Robotics Unit
CES	composite emotional states
CO	Cognitive
CRM	Consciousness Reference Model
DS	Data-flow Selection
DV	Desirability Vector
EAA	Emotional Agent-Architecture
EBA	Emotion-based Architecture
ERM	Emotional Reference Model
ESM	Emotional State Model
GToC	General Theory of Consciousness
LO	Locality
NE	Network
NRM	Non-Reflexive Mechanism
OSI RM	Open System Interconnection Reference Model
PW	Physical World
RE	Representation
RM	Reflexive Mechanism
TR	Transmission

List of Fundamentals

Assumptions

Definitions

Statements

Chapter 1

Introduction

More than some other people, I often find myself being absent-minded. Things in my direct proximity pass me by and I sometimes don't even seem to be conscious of my own thought. Sometimes even, I find myself for a short while devoid of thought. Apparently a link in a chain of processes has been severed somewhere, because normally I walk around trying to see and explore as much as possible from the world around me. Why am I sometimes devoid of thought? In what stage of the processes normally leading to my consciousness does that happen? Are there stages?

In my search for my own process of consciousness I have stumbled upon many ideas by many people. Each of these ideas usually carries the same global structure provided that consciousness isn't considered to be some sort of divine intervention. These ideas then consider an implementation of consciousness, where they fail to identify the global structure first. It's to my opinion not always the best option to go into the lower processes of consciousness. Instead it would be better to first describe a more global idea. Most of the time the global idea most people agree on will do just fine for predicting different outcomes of processes and experiments.[1]

I haven't been able to find an abstract model which describes consciousness as far down as possible without ending up in a discussion as to what the lowest level of processes should do. When the lower level processes are discussed, the way the processes at the lowest level should presumably work then obscures the higher abstract processes. In terms of the global architecture of consciousness the way the processes at the lowest level work really doesn't matter that much.

What also happens is that the global architecture of consciousness is only split into too global sub-processes, instead of clearly defined smaller stages. Now each of the still global stages can still be divided into smaller stages without starting a discussion on the implementation of consciousness. This way not all the behavioural aspects of all the functionality is present in the model, even though it really should be.

An intermediate solution has to be found. The smaller stages should be as small as possible, but not thus small that it may be built differently as well. This means that the model should describe the behaviour of the smallest of processes that may lead to consciousness properly.

[1] For a review of some of the literature which I have read and the reasons why I came to the conclusions named here, I refer to chapter 4.

One needs to acknowledge that it doesn't matter how certain behaviour is generated in certain functionality. If you know what the outcome should be, does it even matter how you get there? If you want to go from Amsterdam to Paris, except for efficiency differences, will it matter in the end if you go by plane or get there by car? In the end needed or expected efficiency should be the only thing determining which implementation is chosen. But before efficiency may be discussed we first need to understand the general architecture. Efficiency may be seen as an embodiment problem and not a problem of mind.

1.1 The Goals of Research

In this thesis I propose an abstract model describing the basic processes needed for consciousness to arrive. This model defines the processes as far down as possible (without making guesses) not taking into account the different architectures that may be used to constitute a conscious being. The architecture that makes up a being then constitutes a certain number of these processes. These processes need to be present either in an implicit or an explicit form. These processes are defined independent of the substrate.

The model can describe any type of being. The different opinions about what the processes at the lowest level of abstraction should be made up of thus aren't modeled. They should be implemented defining the internal functionality of each of the processes for experiments by the people who want to use that functionality. Again, the implementation will just make a difference in efficiency.

The model defines a clear basis that people of different research areas may use to describe and display their ideas. The interaction that may come forth from this should stimulate better ideas and better results in research.

It's important to understand the behaviour of the functionality better. This will also lead to a better understanding of the necessities that may arise with each of the different layers properties.

The model may form a basis of proof to some people that they are wrong or to others it may help show that they are right. There are many speculations about what the layers should eventually do of which many without scientific proof and often even in denial of science. Using the model it's now possible to provide an interdisciplinary discussion of all of the ideas emergent in society.

The model provides a better model for interaction which may be used in more general applications. This way applications should behave more life-like and be less (or sometimes more) annoying to people than they usually are. What's meant here is a discussion of things like joy, pain, humour and many more of the interesting facets of for instance human-nature. The model will also serve as a start to explore these interesting research areas which will be largely captured under the heading of emotions.

1.2 Approach

To make sure that all of the goals are reached as fast as possible I should first look at what has already been done. What ideas and suggestions regarding the process of consciousness have already been displayed? What is the current

state-of-the-art? And most importantly: how can I fit this into a single abstract model?

After having collected these ideas and ordered them globally some thought has to be spent on how to model them on an abstract level. To do this I will have to find a proper way of modeling them, where the rules for the model are as complete as possible in their design. One of the major set-backs in modeling is the lack of completeness in the way to describe a model. Not to mention to find the right abstraction level.

After having decided how to describe the model, it's time to think of the actual description of the model. This will need to consist of the model itself as well as practical examples found in everyday experience to provide for a sufficient foundation for the model as I propose it. The model should be built on a sufficiently high abstraction level. This means abstracting things of which we know they should happen, but don't exactly know how they happen or may happen differently under different circumstances.

Having constructed the model of the processes that should lead to consciousness, the model will have certain implications regarding everyday behaviour of conscious beings. Since these beings are part of the world and interact with the world, they may eventually also interact with each other. It's possible to describe different interaction mechanisms between and the effects that these may have on conscious beings. The effects are for instance feelings, or sentiments as you might also call them, which evolved in conscious beings.

1.3 The Structure of the Thesis

The documentation of this research will be structured as follows. Chapter 2 proposes a division of the different processes into layers. This defines on a very high level the relation between and the function of the separate layers. Chapter 3 discusses the choice of modeling language. Chapter 4 provides a small overview of literature which influenced my perspective of researching consciousness. In most cases however, the literature will just be cited along with my research. Chapter 5 describes the different layers in more detail, providing the necessary argumentation to why each layer is modeled the way it is. Chapter 6 then discusses the implications of the described models in everyday behaviour of conscious beings. Chapter 7 discusses why the model is correct. Finally I will have to draw my conclusions and write my recommendations to clarify what has been and should still be done.

1.4 On Reading

The definitions given in the thesis are given in such a manner that they can be represented as logical rules under certain assumptions. With each definition it's of course possible to ask further and further referring to its meaning until there are no words left.

The idea thus is that it should be possible to state definitions clear enough to identify necessary symbols in the working of processes, but not to define it further as to what those symbols are made up of. This is also why it's the General Theory of Consciousness (GToC): it describes what needs to be

implemented, but not how to implement it.

To my opinion there is not just one solution. Arguing about what *the right solution to consciousness* is is probably the thing that has slowed down the coming to an actual solution the most. Instead of working towards all solutions, philosophers and scientists alike are mostly arguing amongst each other instead of working together to find a general solution. This needs to change.

The report focuses on the removing of any ambiguity from the definitions. Just as well the definitions propose a closing theory without contradictions. The proposed theory is derived and in accordance with perceptual reality. My beliefs regarding the way research is conducted are in accordance with Popper.[2] These believes about the truth of theories and statements as well as about (as Popper states it) *demarcation* are supported by the way consciousness works as is expressed in this thesis.

[2](Popper, 2004)

Chapter 2

Introduction to the General Theory of Consciousness

To describe the GToC in this chapter a layered model will be introduced called the Consciousness Reference Model (CRM). This model determines the consequent processes as present in the global process known as consciousness. These consequent processes each will be defined in such a manner that their external behaviour is a clear enough representative for different implementations. To actually derive a working system some implementation still needs to be made. The current model is a flat model, the depth of the model follows from the implementation. The actually implemented model will thus not be flat.[1]

The model I propose constitutes of seven layers and is derived from a networking model, the Open System Interconnection Reference Model (OSI RM). As you can see in figure 2.1 there is a clear correlation between the two models. Just as in the OSI RM the processes in the CRM don't just propagate up, but they also propagate down. Within the CRM there are basically two extremes, represented by the top and the bottom layer. These extremes are respectively the cognitive extreme and the world extreme. Each of the two extremes will have its influence on consciousness. To come to consciousness there will be a definite competition between the two extremes. Chapter 5 will discuss this further.

In the remainder of this chapter I will globally provide a discussion of each of the layers and their functionality. Before I do this however I will need to describe the basic principles behind the OSI RM. Knowing how the OSI RM basically works is sufficient for understanding how the derivative model came into being.

[1]The model as proposed in this thesis is basically a collection of processes which may to some extent run in parallel to each other. It's also possible to implement them fully serially, but this need not be the case. The implementation of the processes could of course also be defined as a sequence of processes which may be inserted in place of a process which needs to be implemented. To keep things clear however it's better to have the CRM as a top-layer model regarding the design of the implementation. Subsequent layers then define the actual implementation that's made. That's why this thesis states that the implementation defines the depth of the model. It's important to keep a clear view of the way the model works. This could be compared with different models implementing the Open System Interconnection Reference Model (Tanenbaum, 2000) on which the GToC is based, but this is beyond the range of my research.

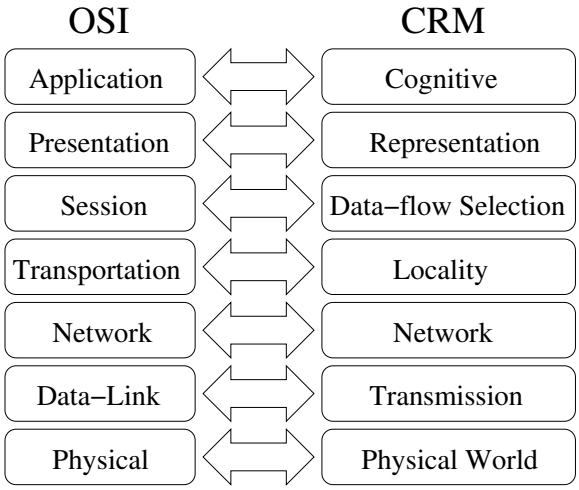

Figure 2.1: The OSI RM and the CRM

After having described the OSI RM it's then possible to describe each of the layers as they should be derived in the CRM. The layers will be discussed bottom-up, because the paper is concerned with a physical description of the processes that should be present in any kind of being to let consciousness emerge. These processes run on top of or within the physical world as we know it.

It's of course possible to start a discussion on whether the explanation of consciousness shouldn't commence at the cognitive level. Philosophical viewpoints are shared at the end of chapter 5 discussing the different *laws* that should hold for a philosophical treatise of consciousness and a physical description of the processes that lead to consciousness. Chapter 5 also refers to the correlation between the two.

2.1 The OSI RM

You most probably wonder why the OSI RM has been chosen to use as a first indication of the actual model that should be implemented. Let's start by explaining this first, after which the OSI RM itself may be discussed.

Both artificial and non-artificial beings depend on certain kinds of networking functionality. This networking functionality is used to form relations which tries to establish the same relations which are present in the world. From these relations then follow certain reactions to these relations. Basically the cognitive image is a representation of the physical world relations.[2] It may be a computer network, or a nerve system. Just as well it may be a functional representation of a network implemented using different kinds of data-storage. For instance within Information Technology one might use actually so-called data-stores.

The OSI RM is one of the possible networking models that are available. There are a lot of other models, but none are as generalised as the OSI RM

[2]With cognitive image here isn't meant the sensation of our experiences, but just the raw establishment of data representing the physical world. Cognitive in view of this thesis means physically representable knowledge. Cognition therefore may be a small part of consciousness, but isn't consciousness itself. Cognition is the ability to learn.

and none are as complete. The OSI RM is complete. This makes sure that it's the most solid foundation to start a discussion upon regarding information exchange. That's pretty much why I chose it as a basis to derive the CRM.

The functionality contained and described in the OSI RM not only describes the basic necessities for computer networks, but also for the cognitive networks needed for consciousness to arrive. In case of cognitive networks the basic ideas of the OSI RM will still have to be cast into a view more appropriate to cognitive sciences.

The OSI RM is a layered model. As you may see in figure 2.1 it has seven layers. The relation between the OSI RM and the CRM was primarily drawn based on basic knowledge of the nerve system. Because of the abstraction level however it now also became possible to apply it to social networks and any other type of consciousness. Chapter 7 describes why each of these processes needs to be present and should occur in the named order. The OSI RM was constructed according to the following five principles[3], which still hold after deriving the CRM:

1. Where a different level of abstraction is needed, a new layer has to be created.

2. Every layer must have a well-defined function.

3. The function of every layer must be chosen keeping in mind the definition of internationally standardised protocols.[4]

4. The boundaries between each of the layers must be chosen thus that the amount of information that has to be transferred using interfaces is as small as possible.

5. The number of layers must be thus large that the different functions don't have to be put together in the same layer, and yet so small that the architecture doesn't become awkwardly big.[5]

In the OSI RM this has led to seven layers. These layers are:

1. **The Physical layer.** This layer is concerned with the transmission of raw bits over a communication channel.

2. **The Data Link layer.** The main task of the data link layer is to transform a transmission facility to a line that looks to the network layer like it's free of undetected transmission-errors.

[3](Tanenbaum, 2000)

[4]In terms of the CRM this will guarantee the generality of the model. This means any implementation which contains the specified functionality is a sufficient implementation. In terms of consciousness this means that any implementation of the layers functionality is sufficient as long as it portrays the correct described behaviour.

[5]In order to keep the basic model comprehensible it shouldn't contain too many layers. The functionality should be spread over different consequent layers in a logical manner. What should be kept in mind however is that some functions should be caught in the same layer, because they are part of the same class of functionality. So each of the layers should clearly identify an abstract class of functionality. This functionality should then still be implemented. Here different implementations form different derivations of the same abstract class.

3. **The Network layer.** The network layer is mainly concerned with determining the subnet that's used within the whole network to pass information.

4. **The Transport layer.** The transport layer accepts data from the session layer and makes sure that it arrives properly at the other side, sometimes by breaking up larger pieces into smaller ones. The transport layer defines the type of transportation services that are offered to the session layer. It also defines which process should receive what.

5. **The Session layer.** The session layer enables users on different machines to create a session together for data transmission. The session layer also arranges the dialogue between each of the participating connections. To make sure that not every process can be busy at the same time, it also provides a so-called token-management. The process that owns the token may perform its critical operations. The other processes have to wait. Another thing the session layer has to provide is synchronisation.

6. **The Presentation layer.** The presentation layer performs tasks that are carried out so frequently that it should provide these functions by default instead of letting the user find the solution to the problems. A good example is the agreed upon encoding of a certain program.

7. **The Application layer.** The application layer contains different protocols that are often necessary for different systems to communicate. Another function of the application layer is file-transfer, where the differences of representation of a file on different systems have to be gotten rid of.

2.2 The CRM

The CRM has seven layers like the OSI RM. These layers are:

1. the Physical World (PW) layer,
2. the Transmission (TR) layer,
3. the Network (NE) layer,
4. the Locality (LO) layer,
5. the Data-flow Selection (DS) layer,
6. the Representation (RE) layer and
7. the Cognitive (CO) layer.

These layers are discussed in this same order.

2.2.1 The Physical World Layer

The lowest layer in the CRM is the PW layer, which forms the basis for all interaction between all the layers that run on top of it. In a sense the PW also seems to contain all the layers that run on top of it. Looking at a single

arrangement of the PW it doesn't actually do this. For the PW to be placed in relation with each of the layers it will also need a certain time-frame to contain the relations within that time-frame.[6] Of the PW it's possible to take a moment in time with an infinitely small time-frame and look at its arrangement. But this will not show all the relations that are currently defined on top of or within the PW, because these relations have a time-factor associated with them. Basically all the layers run on top of the PW, where the PW forms the basis of the implementation. That's why all the layers including the PW layer don't run within, but on top of the PW. They define the relations on top of the PW and not within it.

So the layers are not part of the PW but are associated with it, given a certain time-factor. From a philosophical point of view this is a very important issue. It defines how a process may perceive itself to a certain extent, but cannot predict what its future steps will be because of the lack of possibility to see it all. This defines for instance what Manzotti meant by stating that we can perceive our brain, but cannot perceive the process of consciousness.[7] We can look at moments in time, but we cannot measure consciousness because it's not associated with one particular time-frame.

What the PW basically does is that it carries, receives or emits signals. The signals just propagate until they hit something. They propagate using the relational structures that make up our world. When they hit something they may propagate a new signal. It's a bit like a chain reaction which may or may not end at a certain point in time. Energy in itself is of course just another form of a relational structure. Most probably it should be noted that it's the lowest layer relational structure which may be possible to identify by us. Energy thus could be assumed to be our basis of everything. But this is just an assumption we make here. Each process is made up out of one or more of these signals. An association could be drawn between Manzotti's *onphenes* and these signals, because these signals as I envision them define the relations between parts of the PW.

definition 2.1 (onphene) *An onphene is:*[8]

1. *A physical process*
2. *Corresponds to a phenomenal content*
3. *Is in relation with other entities*

Since my signals contain a large amount of ambiguity regarding the interpretation of the word *signal* itself, the *onphene* definition is here and now assumed to be the representative of the different processes leading to consciousness. In order to make a clear distinction between the different subsequent processes I've also had to introduce quite a lot of definitions myself. All these are to a more or lesser extent specialisations of the *onphene*. One process can of course be made

[6] The PW layer here isn't the same as the PW. The PW layer identifies part of the functionality that should be part of the processes running on top of the PW to let consciousness emerge. For instance a world of zero degrees Kelvin doesn't process any signals. No changes are theoretically made to the world by signals which run through it. Of course this isn't wholly true since the world will be warmed up, but it still serves as a good example.

[7] (Manzotti, 2003b)

[8] (Manzotti, 2003a)

up of multiple smaller processes, so the *process* and the *signals* as I used them are both instantiations of respectively larger and smaller *onphenes*.

Chapter 5 will contain my own necessary definitions to define the sought for processes. The definitions will not be concerned with the term *signal*. To associate them a bit with feelings we have, I will use the term *signal* widely in explaining the association that we should have with each of the definitions. This is regardless of the ambiguity of its meaning. In the end we shouldn't be speaking about signals anymore, but just about the terminology as stated in the definitions.

To some people it may not be particularly clear what is meant by the *phenomenal content* which is contained in the *onphene*. It seemed in order to add a small discussion regarding this issue here. *Phenomenal content* basically is the actual thing that we experience. For example, in case of the colour red *phenomenal content* is not just the determining that something is red, but the actual experience of red itself. For instance I could associate constants in a computer with the colour red, but this would arguably not provide the computer with the experience of red as we have it.[9]

I asked Manzotti if this idea was correct and he agreed. He also gave the following response which poses an interesting question which may need answering in the future:

> *"I would say that the phenomenal content is what is phenomenally perceived when a conscious subject has an experience of something. When we open our eyes we have an experience of something. This something is the phenomenal content of our experience. Of course this does not mean any commitment to a dualistic point of view. All options are still open. Of course I believe that there must be a physical counterpart of this phenomenal content. The viable options are (I think):*
>
> 1. *some pure mental content different from physical reality (I would personally reject this very substance dualistic case);*
> 2. *the neural activity in the brain (I reject this option since it seems to me a kind of physicalistic dualism);*
> 3. *the external object or event (fine but it has practical and logical problem);*
> 4. *the whole process engaged from the external environment up to the brain (It is my favourite choice)."*

Since the *phenomenal content* can be perceived itself, in the CRM this is induced into the PW by the processes of the CRM. This way the phenomenal content becomes an object or event on itself which is theoretically external to the being. This means that the phenomenal content is induced into the PW layer and is then again received by the rest of the processes which make up the being. Again should be noted that cognition is only the part of consciousness which is occupied with learning or a physical representation of the knowledge base. Just

[9]The actual experience isn't defined in this thesis, nor does it need to be to discuss consciousness. See also (Hobo, 2004b, Section 2.3). This also discusses *qualia* and their relations to the CRM and consciousness. Qualia as we have them are then just yet another implementation.

as well it now becomes more clear that even regarding phenomenal content it may well be possible to have different implementations. These implementations then should lead to the same behaviour.[10]

Manzotti proposed four options in the above response. Since I'm dealing with a science of consciousness and not a religion I personally have to reject the first option.[11] But even for people who don't it's possible to fit this into my model. All you need to do is declare a receptor for us that transfers the experience from physical reality to this non-physical reality. The main model as described in chapter 5 and its consequences will remain correct for all options.

2.2.2 The Transmission Layer

When a signal containing information progresses, the information contained by the signal doesn't necessarily change with the changing of the physical representation of the signal. Usually the information is transmitted further by using a different kind of encoding of its content for each medium which is used to transfer the content. For instance one might have a light-receptor on one side (a camera) and a light-emitter (a television) on the other side with an electrical cable in between. The wavelength of the light can be encoded by the receptor into an electrical signal after which the signal is transmitted via the wire to the emitter. The emitter then may decode the signal and emit the specified wavelength. The television image here takes the role of an actuator because it casts a signal out of a clearly defined system. This of course is just a small example. As you may have noticed a camera and a television don't make up a conscious being. Although the process is dynamical in execution, it's still a static process in time because the same process is repeated over and over again. The process of consciousness however is a continually changing dynamical process.

A conscious being on the one hand receives signals from the PW, but it may also emit signals to the PW. The receptors receive (not perceive) signals from the PW. This means that for instance a ray of light may hit your retina. The emitters emit signals to the PW. This may for instance mean stimulating muscles to contract. Emitting a signal doesn't mean for instance pushing a boulder, because this is an act within the world itself. The pushing action and sensation come forth from indirect stimuli by the actuators to the receptors via the muscles and the boulder. A signal from an emitter is just a stimulus, not the result of the stimulus. The result of the stimulus is a sequence of one or more processes itself.[12]

What we now have is the receptors and emitters which together constitute interfaces between the PW layer and the TR layer for signals. In receiving or emitting signals it's important to reduce the noise. This means that the TR layer may only propagate a signal if the signal exceeds a certain potential. This way the TR layer should make sure that all collections of signals propagate

[10] (Hobo, 2004b, Section 2.3)

[11] Science and religion differ based on the fact that science is derived based on observation and doesn't lead to contradictions. When supposed science does lead to contradictions it is in that very instance reduced to religion.

[12] Some people wonder what place a thought has in consciousness. A thought is in itself a chain of processes. Each step within this chain will induce certain signals into the PW which will then be perceived again piece by piece. The whole thought is then the chain of these signals which in turn are received again by the TR layer and processed up.

properly. It may not be fully clear what *noise* means. This is captured in definition 2.2.

definition 2.2 (noise) *Noise is all irrelevant data.*

What should here also be noted that when looking at a neuron, the outside of the axon might be considered to be an emitter. The outside of the dendrites then are the receptors. The signals that are transferred between different neurons are propagated through the space in between the neurons. The space in between the neurons then is the PW layer and the surface of the axons and dendrites are the emitters and receptors. This way a network of a single conscious being is formed.[13]

2.2.3 The Network Layer

Signals may be transmitted via different paths to different endpoints. The NE layer should determine what the best route is, as well as translate signals in such a way that all receiving parts of the network properly understand what was meant. That may for instance mean from a biological point of view that electrical signals will induce the release of certain chemicals, which in turn induce other processes. The route may then be determined by the receptiveness of a certain connection regarding the signal that's produced. For instance when the CRM is used modeling single neurons these processes are formed by whether two neurons share connection points using their axons and dendrites.

2.2.4 The Locality Layer

The LO layer has to clearly define what part of the subject emitted or should receive a certain signal as well as provide a means for different parts to communicate. This also means that it should make sure that signals should be propagated to multiple recipients when they should need it. For instance the muscles in the stomach should in some cases be contracted in unison. Now each of the emitters for the muscles should receive a signal that they should then emit a stimulus to the muscles at approximately the same time.

The LO layer should also make sure that the synchronisation of certain sequences of signals should be in order. This may be done using for instance a certain time-stamp with each of the emitted signals. It may just as well rely on the proper timing in emitting different sequences of signals.[14]

What it does is defining the fact that certain LO aspects are and should be present. Other than that it just passes the signal to the next layer.[15]

[13] As will hopefully become clearer in the rest of the thesis, the whole of a conscious being is made up of multiple instantiations of the model. These work together in a networked fashion. The conscious being may be a single being or it may for instance be a group of beings. The group then forms one conscious being.

[14] In the past I here made the mistake of defining the implementation. It of course doesn't matter whether the layer is explicitly or implicitly modeled.

[15] The LO layer may be placed in relation to the NE by stating that in order to define the LO first a topography should be present. Without a topography it's not possible to discuss where signals originated from. The topography is captured by the implicit or explicit instantiation of the NE.

2.2.5 The Data-flow Selection Layer

To be able to perceive things certain abstraction mechanisms will have to be in place. Just as with the TR layer which filters out noise, there should be a pre-selection regarding the signals that should be listened to and which not.

For instance when we listen to someone, we focus on that one person. This is something that we have to commence in actively by choice or *desire*. If we receive too many stimuli by for instance a lot of voices in a very crowded area, this will make sure that it's impossible for us to single out that one person we're speaking to. So the DS layer forms a process that requires to some extent active choice and on the other hand it may also be influenced by the signal potential that we receive.

Another example of signal inhibition that surpasses choice is messages contained in for instance films. These may be caught in a single frame so we can't perceive it actively, but it's then often perceived subconsciously and influences our behaviour. We perceive them because we chose to watch the film, but the frame is too short for us to truly perceive it. Sometimes people will actively perceive it because they are really fast. Most people may deal with such signals in dreams during night or day.

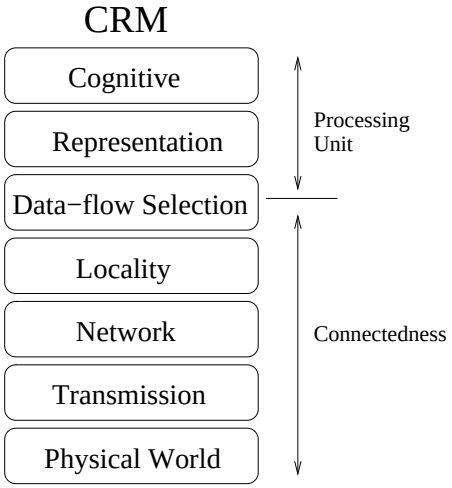

Figure 2.2: The separation between connectedness and processing units within the CRM

The DS layer basically forms a boundary between the connectedness and the processing units. Connectedness is basically connections between different processing units as well as between the processing units and the PW. This is shown in figure 2.2.

2.2.6 The Representation Layer

For the mental processes to be able to do something with the received signals a clear representation (RE) should be built to be able to work with them. The RE should be built by the mental processes themselves, since it's directly linked with the functionality of the mental processes. This is done in the RE layer.

This RE will just be a RE of the gathered input signals. Just as well it serves to transmit output signals. These may be processed so they can be emitted into the PW layer.

Just as well the RE layer may build a RE of the eventual outcome of certain possibilities of signals that it may produce. For instance Manzotti gives different examples regarding shape-mapping.[16] Manzotti explains it a bit differently, but I explain it as follows. One of the examples was the mapping of a cross onto a matrix as displayed in figure 2.3. When the relation between the different numbers in the matrix is found, it's possible for a subject (if the shape *cross* is known to the subject) to project the shape of the cross onto the numbers that form the relation. Now the subject may apply the term *cross* to the ordering of the numbers in a new reference frame. This applying of terms to a perceived image of the PW is of course done in the CO layer where information is resolved after which a new RE is built including the found solution.

$$\begin{matrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{matrix}$$

Figure 2.3: A 5×5-matrix containing a cross of ones

The image of the PW that is built should be defined as a RE of the collection of received and transmitted signals. The RE could for instance just be a collection of bits or some other functional representation. The interpretation of the RE is not part of the RE layer, so we shouldn't try to add some kind of feeling that we have with RE to this layer. The perception is partially induced by the CO layer.

Although it may look like the afore-mentioned RE of a cross is only built within the mental processes, this will not be so with the model I propose. The RE itself will need to be built using so-called actuators (which are emitted by the emitters) and is then again perceived by us. So the cross is physically expressed into the PW and then perceived again combined with the original picture. Thus the process of perception is a continuous propagation of signals up and down through all of the layers.

2.2.7 The Cognitive Layer

The CO layer is the final stage in the process of perception and the first stage in the process of exploration. Everything that is processed up from a lower layer should be resolved here. Just as well all decisions made regarding the *desire to influence the world* are processed down.[17] This means that this layer should be equipped with the recognition of objects, relations between objects

[16] (Manzotti, 2003b)

[17] Note that it's a desire to influence the world. This doesn't mean that it's a felt desire. This means that the generated desire may lead to influencing the world but it sometimes can't. Suppose for instance that a muscle is temporarily sedated, which means that it doesn't receive any signals. The brains may now send stimuli to this muscle but they do not actually arrive. Here the desired effect isn't reached. In case of us this may of course also lead to a felt desire, but this is, although related, a different set of processes.

and the relational flow or (to put it in other words) the change of relations between objects. It should also be able to determine what it does and doesn't know. Because if you don't know that you don't know, why should you explore? Through exploration a being also gains experience which alters the CO layer. This is of course part of learning. More about these CO processes in chapter 5.

2.3 Interpreting the Model

During the writing of this thesis it became clear that to some people when learning to work with the model it's justifiably hard to interpret the model as proposed. In order to aid these people in their effort here follow two possible interpretations of the model. First follows the interpretation of nerve systems. Then follows the interpretation of a social NE. There may of course be many more.[18]

2.3.1 Nerve Systems

The basic structure can be identified as our nerve system. But how to interpret each of the parts of the nerve system? The nerve system itself is a large NE. If we take one piece of the NE like a neuron it's then possible to apply the previously described layers. How this happens is now described in the rest of this section.

The PW layer constitutes of everything that happens between each of the small pieces that make up our nerve system as well as the world exterior to the nerve system.[19] The signals that for instance are passed between the neurons are passed from axons to dendrites. Between these axons and dendrites there's this infinitely small piece of nothing which may be used for the propagation of signals. These signals will be contained in electrical currents and chemical reactions. Of course the senses of a being form the connection with the outside PW.

The TR layer consists of the parts of the axons and dendrites functionality in emitting and receiving these signals. Just as well the world external to the nerve system will pass its signals onto the NE by our sensory organs.

The NE basically is made up of all the connections that are there between axons and dendrites as well as between our senses and the nerve system. The senses here then serve as TR points to and from the world external to the nerve system.

The LO layer is associated with the specific time and location of a certain axon, dendrite or sense where it either sends or receives a certain signal.

The DS layer determines to which extent some instance of a neuron listens to other neurons as well as the senses. In case of neural NEs in classical AI this can to some extent be associated with the weights and existence of connections

[18] Although I conceived of the model, I too had to learn to work with the model and am still learning. I too have the problem of thinking in types of implementations of the model too much sometimes.

[19] This doesn't mean that it's a general theory of everything. This means partly that there's room between each of the neurons between which they pass signals. Just as well it means that the exterior of the nerve system is connected to the PW from which it may receive signals. These signals also originate from space. So the PW layer defines the space through which different signals progress.

between all the neurons as well as the senses. The only difference is that in classical neural NEs the neural NEs will have constant weights at some point in time and in case of the neural NEs we have in our nerve-system the structure is fully dynamical.

The RE layer basically determines how the signals are passed. Are they passed as electrical currents, chemicals or maybe integers? So basically the RE is made up by the exterior view of the signal which isn't the same as our *perceived* RE. It's the physical RE as determined by for instance the science of physics. Of course signals could also be passed as floating point numbers, but this is just another physical RE.

The CO layer is then the CO function which gathers all the inputs and determines a certain output. This is determined by the internal make-up of a neuron. Cognition means learning, so the CO layer contains the learning structures.

2.3.2 Social Networks

As most people have noticed social NEs in society can also lead to forms of consciousness or single-mindedness. This can usually be determined by waves of emotions just as a signal perceived by the senses may propagate as a wave through neural structures within our brain. So how can we interpret each of the layers in case of social NEs and the being of society?

To start with there is again the PW layer. This PW layer is used to transmit letters, the daily news, radio and television signals, but also for instance speech and a whole lot of other signals. This permits parts of the social NE to communicate with other parts of the social NE in a more or less direct manner.

The TR layer is then made up of the places where each of these messages are emitted and received. This may be our mouth, but just as well a postal box, a broadcast station in association with a television set or maybe our ears.

The NE is then made up by all the relations between different parts of the social NE. These relations are then made up by postal services, the Internet or maybe a direct meeting within a room where people can communicate more directly.

The LO layer is made up by the time and place where for instance a postcard is sent. Just as well it may be before the door of room 305 in a hotel where a person yells to a person at the door of room 297. Of course it may also mean that it's a combination of a television studio and a television set as previously mentioned.

The DS layer is then made up by the tuning in to a certain signal by a person or not. A person may decide not to listen or maybe turn off the television set. Just as well a person may throw out anything from its postal messages which is wrapped in plastic and is printed on bright shiny paper.

The RE layer determines what the message looks like. It may be writing on the postcard or maybe it's the programming of the nine-o'-clock news.

The CO layer is then the person itself which handles the selected input and chooses to form and express an opinion or not.

This basically transposes *the Chinese Room* to a new perspective where the person in the room may arguably not be conscious of its task. On the other hand the person may now be part of a larger social structure. For instance if the

person is part of a message-filtering system but doesn't know what its filtering, it does influence society. It just doesn't understand its function.

Chapter 3

The Choice of Modeling Language

Before continuing in deriving the actual GToC, it should be well established which modeling language will be used in deriving the model. To start with, it's important to look at the requirements of such a modeling language. Based on the requirements it's possible to discuss the different modeling languages. Then one of the modeling languages has to be chosen to actually derive the model with. Finally a short guide is given to the modeling language.

3.1 The Requirements

There are a few things that need to be considered in choosing a certain modeling language.

- It should be easy to understand and thus visually appealing.
- It should be complete and well established within science for dependability.[1]
- Although not everyone can be familiar with the models that are used, the models together with the provided explanation should be clear to everyone. This means that the thesis should provide plenty of information, in correlation with what is modeled.
- The models shouldn't be insufficient where the explanations would have to make up for the insufficiencies.
- The model should adhere to formal semantics, where the semantics should be as generally comprehensible as possible.[2]

[1] This means that it should be a complete non-contradictory closed system.

[2] These formal semantics may be of any kind. For instance ancient Egyptian languages adhere to certain formal semantics, even though it isn't comparable to any language we use.

3.2 The Modeling Languages

There are quite a lot of modeling languages. I'll divide them into three categories: plain text, formal languages and graphical models.

1. **Plain text.** Although this can contain everything that's needed, it will take a lot more careful reading than for instance a graphical representation of a model. Just text will not do for another reason also. In using a plain text model in order to implement it in any way you would still have to a build model of the text model afterwards, suitable for implementation.

2. **Formal languages.** An example is the language of Z.[3] Again these languages can contain every kind of information that's needed, but they are not visually attractive and contain a lot of mathematical expressions. With or without having mathematical feeling, it should always be as clear as possible what is modeled. In case of formal languages this isn't possible. Even if you have a clear conception of mathematics, it's still easy to lose track of what's happening. You still have to build a mental image of what's written. This image should be provided as much as possible by the modeling language as used.

3. **Graphical models.** A good attempt at a visually appealing language is made by Manzotti[4] by his graphical representation of onphenes. Its comprehensibility doesn't suffer without textual explanation, under the assumption that each of his onphenes receive proper names. Unfortunately it isn't complete and most probably still under heavy construction. This could be compensated for by adding your own rules to the language, but this will give rise to a rather large probability of error. Fortunately in models that closely resemble Manzotti's work, this has already been done. These models have been in use for a longer period of time and they have proven themselves worthy and dependable. The states in correlation with Manzotti's work may then be referred to as being *onphenes*. For the definition of an *onphene* I refer to definition 2.1. The modeling language that I'm after is the one captured in the action relation diagrams for describing causality as proposed by Vissers et al.[5] To summarise the modeling language's most important properties:

   - The language is visually appealing.
   - The language contains conjunction, disjunction, recursion, attributes and much more.
   - The language also contains the possibility to model the operations that take place during each of the steps in the relational structure plus the causality conditions of the relations (uncertainty, conditions under which things may happen, etc.).
   - It's easy to make a clear distinction between each of the layers and it's easy to show the interaction between a certain set of layers.

[3](Jacky, 1997)
[4](Manzotti, 2003b)
[5](Vissers et al., 2002)

3.3 Conclusion

I hereby propose to use the action relation diagrams for describing causality as proposed by Vissers et al. Quite clearly these are well suited to describe the *onphenes* as proposed by Manzotti. Since the CRM is an instantiation of many of these onphenes this is very important. The specific language adheres to more generally and intuitively known modeling techniques, contrary to for instance UML. It isn't a matter of whether it can't be done in another language, it's a matter of which language is most appealing.

3.4 Short Guide

When going through chapter 5 there are a lot of figures. There's a subdivision in layers and onphenes, where larger onphenes may be subdivided into smaller onphenes. Here follows a short guide to what the graphical representations of these onphenes mean.

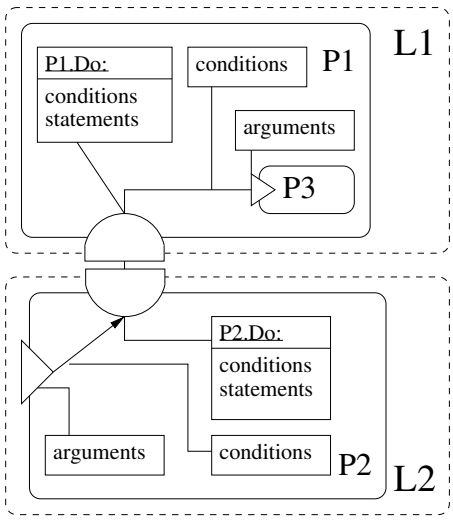

Figure 3.1: Reference figure

When looking at figure 3.1 a few things may be noticed. The layers are indicated by the dashed boxes and have received the name *L1* and *L2*. *L2* contains a process *P2* with an *entry point*. The entry point is where a process is initiated. This *entry point* is indicated by a triangle. The *entry point* may receive certain *arguments*. These *arguments* basically are the input values for a process. These *arguments* receive a name at this entry point where they are declared for the first time. When *arguments* are declared for the first time they also receive a *type*. Only when referenced after declaration can the *type declaration* be omitted in specifying the *arguments*. A *type declaration* of an *argument* looks like *'argument: type'*. The construction of this kind is needed to properly pass values between consequent processes.

Within *P2* an arrow indicates the order of *change of relations*, where each change is commenced by an *action*. This *change of relations* may then receive

certain *conditions*. If a *condition* doesn't hold the *change of relations* doesn't take place.

When the *change of relations* takes place this may be indicated by an *interaction point*. This is indicated by a semi-circle which indicates an *action* is shared by multiple processes. Here the *action* is related to a certain process (which in turn is a kind of function). This process which is indicated by the *action* then changes the relations. An *interaction point* receives for each process *conditions* and *statements*. These need to be executed in the specified order. When a condition doesn't hold the process taking place at an *interaction point* ends and the more global process may continue.

P1 also contains an *interaction point* which cannot continue before the *interaction point* at *P2* finishes (and vice-versa). When the *conditions* of the *change of relations* in *P1* hold it calls a process *P3*.

In calling the process *P3* certain arguments may be passed at its *entry point*. *P3* may then use these arguments in its own process.

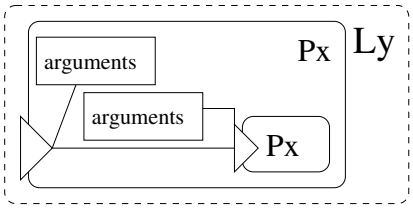

Figure 3.2: Reference figure

A process may also use *recursion* by calling itself. In figure 3.2 the process *Px* in layer *Ly* calls itself. It receives *arguments* at its *entry point* which it passes on to an instance of itself.

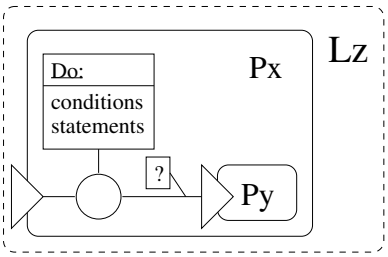

Figure 3.3: Reference figure

When an *action* only takes place within one specific process, this is indicated by a circle. This is illustrated by figure 3.3. This circle receives certain *conditions* and *arguments*. These are again executed in the specified order. When a *condition* doesn't hold, the *action* is finished and the process may continue. This may mean that there are possibly synchronised and shared actions. But this largely depends on the possibilities of the implementation. (The implementation differs greatly based on whether parallelism may be used or only sequential processes are allowed.)

What may also be noticed is that sometimes you don't know whether a certain *change in relations* takes place, regardless of any conditions. In such a

case, the uncertainty may be indicated in the form of an *uncertainty attribute*. This is represented by a question mark.

Chapter 4

Previous Research

In the past many scientists from more and less philosophical backgrounds have been concerned with our being conscious. My supervisors quite understandably argued that it wouldn't do these people justice not to list their research that I've reviewed to some extent. I agree with my supervisors. What here follows is a summary of the literature research which I've done during the writing of my thesis on the GToC. What should be noted is that the idea as portrayed in chapter 2 came first based on a hunch. Only then I committed myself to searching for the material to substantiate this notion of a GToC. Many authors and their research which have been listed in the rest of the text, haven't been listed here. Writing everything down regarding my literature research would have been a thesis of its own and would have left me with no space to display my own ideas. I have included proper references.

4.1 Riccardo Manzotti

When I first came up with the idea of the GToC in Zürich during my practical training at the Artificial Intelligence Lab in Zürich (AI Lab), Gabriel Gómez said that I should ask Manzotti to be my supervisor. Although Manzotti was a bit busy at the time, he did agree on aiding me when I had any questions. From Gabriel I first received a paper on the notion of *onphenes* by Manzotti. The one I received was an early draft of a paper for use within the project ADAPT at the AI Lab. Most importantly this paper showed me how the processes occurrent in the larger process of consciousness should be properly defined. The *onphene* definition[1] spared me some valuable time in explaining people what I meant by stating for me what such a process should contain.

Reading Manzotti's paper quite naturally led to me reading more of his work, namely his Ph.D. thesis. This basically described a possible implementation of the CRM and served to show how the CRM formed a generalisation of this model[2]. Most fortunately everything that was needed was modeled either explicitly or could be stated as a property of the CO layer.

[1] Definition 2.1
[2] Section 5.8

4.2 David J. Chalmers

When communicating with Manzotti, Manzotti sent me a list of books which I could read. It was a list of three books of which the first one was indicated to be especially important. This book was the book by Chalmers, "The Conscious Mind".[3] This specific book was concerned with within what context the search for a fundamental theory of consciousness should be placed. His book focuses for a large part on *phenomenal content*. Since *phenomenal content* is enclosed in the *onphene* there was no need for my model to justify this *phenomenal content*. This prevented my model from being caught in for instance metaphysical explanations obscuring the general process of consciousness.

Personally I don't believe that there is some metaphysical aspect to consciousness. Since the model that I propose leads to certain behaviour of a being like us that claims that it has *phenomenal content*, any other implementation which follows this guideline will do so also when the complexity of our person is reached. The *phenomenal content* now becomes a claim rather than an entity which is part of the behaviour of being.

I have only read parts of Chalmers' book because of the lack of direct relevance to what I've done here, so I have to deeply apologise for not reading all of it. It's just that *phenomenology* can to my opinion be only better understood when we have more basic methods of describing conscious behaviour into which specialisations can be fitted.

Apparently I now raise a small contradiction. I state that to my opinion the *phenomenal content* comes forth from the model. But wasn't it part of the *onphene* definition? Yes it is, but the *phenomenal content* might to my opinion be encoded in an almost infinite number of physical values instead of metaphysical ones. This then leads to us having the experience of the phenomenal content. This experience to my opinion is just another physical process. All metaphysical theories should to my opinion be abandoned, because this would then lead to the whole of physics being a metaphysical experience. This would basically mean that physics doesn't necessarily exist at all. When there's a real chance that physics doesn't exist, why even bother about our existence, because we most probably then do not exist ourselves. For some reason I get the hunch that this isn't true.

Chalmers has chosen the viewpoint that also metaphysics should be discussed as an eligible option. My model does not contain my opinion that metaphysics should be abandoned. It is a general model and may thus be applied in any field, regardless of my philosophical opinions regarding metaphysics.

4.3 Karl Popper

A book which I found very reassuring was a book by Popper called "The Logic of Scientific Discovery".[4] I was put on the trail of Popper's work by Anton Nijholt because he found that my method complied to the *there is no one true method of science* idea uttered by Popper. Not only did it support my apparently maybe somewhat unorthodox method of doing my research, it also supported my idea of knowledge and decision making as expressed in this thesis.

[3](Chalmers, 1996)
[4](Popper, 2004)

On the one hand it supports that we can never know something for sure except that some things may be falsifiable by showing the opposite and even then we just have an estimate of what we see. Although the chances are rather high that we see what we see, there are many psychedelic examples of the fact we cannot be sure.

Just as well, although the book disfavours logical induction as a method of proof, it does engage the thought of growth of the possibility that some prediction you make is right based on succession of estimates. This supports my idea that an estimate of the solution to more complex problems can be made by applying in succession smaller estimating processes.

Popper was a great enthusiast for the way Einstein [5] performed his research. The structure of this thesis can somewhat be compared to his paper, which of course says absolutely nothing about the contents of the thesis. I can only mention that my research method follows the same method of explanation as does Einstein's.

4.4 Aaron Sloman

I read a paper part of a collection of papers by multiple authors, this specific paper being written by Aaron Sloman.[6] This paper could be mentioned because it gives rise to the complexity problem of consciousness. I say *could* because it supports my thesis and what I wish to establish. But the idea I had was already there before reading the paper and so was the approach. So the existence of this paper has made me feel more sure about my cause, but it hasn't noticeably influenced the actual process of writing the thesis. The only exception of course being this section.

The paper identifies different layers of processes each adding a layer to the complexity of being in association with emotions. It also makes a distinction between reactive and deliberate processes. I've made a bit more of a distinction between reflexes and planning, but there are of course also processes which are reactive which may grow more sophisticated in due time. This may be related to section 5.10 by describing the inhibition of processes.

In favour of the opinion of Sloman, as stated in a summary of he made of one of his lectures[7], I've done two things. On the one hand I restricted myself to describing an architectural constraint which may lead to consciousness. On the other I've refrained from even remotely thinking about stating that one being is fully conscious and the other quite clearly isn't. My general opinion is that consciousness has to grow through evolution and even within a being during its lifetime. This means that consciousness comes in sizes. To have a size a being of course needs to be there. My model imposes the constraints on what a being should possess to have any kind of size regarding consciousness.

To some extent according to Sloman's lecture I can rest assured regarding the foundation of my model because I did not only depend on first hand knowledge of consciousness. To quite a large extent we are taught even in high school what the general make-up of us human beings as well as dogs, flies and what not is. In order to properly design a general theory in university we learn many

[5] (Einstein, 1952)
[6] (Sloman, 2000)
[7] (Sloman, 1996)

abstraction mechanisms. These grant us the opportunity to leave the content for what it is and focus on exterior behaviour to which content may then be added. Sloman states clearly in his lecture that we should not focus on implementation issues. Or, in his own words:

> *"So, pontificating about whether machines can or cannot be conscious or about whether consciousness does or does not need quantum gravity engines, seems to me to be just silly. There is far too much research still to be done, and then instead of asking a few big, but empty and largely unanswerable, questions about consciousness we can ask, and perhaps answer, a large number of smaller, fascinating questions about all the myriad components of the cluster."*

This goes against the work of Roger Penrose.[8]

After reading Penrose's work it to my opinion becomes clear how easy it is to mystify clear scientific work in favour of almost absurd religious-like ideas. For instance quantum mechanics is subjected to multiple interpretations of doubtable scientific value. Although this also happens within Chalmers' work, Chalmers just discusses for instance the so-called Everett interpretation of quantum mechanics. This interpretation basically means that everything happens, and what we see happening is just one of the truly infinite amount of possibilities which may happen. These possibilities are then spread through the dimensions.

Mystification doesn't lead to an answer, whereas simple mechanisms do. Why then conjure up some complex plan without actually thinking of a question which can be directly related to the problem of consciousness? Let alone why provide an answer to any question? If none of these two issues are addressed, what is? The question is what the basic processes are that may be identified which lead to consciousness. What gives rise to the processes doesn't really matter that much, as long as they are there. Penrose only seems to address the lowest of possible processes and seems to want to see consciousness emerge from there in some kind of strange mystified element. Consciousness isn't an element to perceive, but it's a process to endure.

To some extent Sloman names the mechanisms which have been mentioned in *appendix I*.[9] For some reason his discussion seems to end there. The mechanisms as I describe them, according to Sloman should be organised in layers. Since networking topologies and behaviour may be applied to any kind of consciousness I prefer to propose a networked architecture as I have done. A layered architecture would suggest that the higher level of control always depends on the lower level of control. The lower level of control does guarantee some form of survival in the beginning of a being's life. The higher levels may then specialise certain areas of control to fit the environment's specific requirements. But in some cases I do not agree. I think it's also possible that some higher control without any pre-programming may be attained by first learning to solve simple problems and then specialising into solving more complex problems. Maybe in Sloman's view this was what he felt to be right and should thus be the architecture. The more complex problems are made up of a stack of smaller problems, so these do behave according to a layered architecture. But to my opinion this is the architecture of the problem and not the architecture of the solution.

[8] (Penrose, 1994)
[9] (Hobo, 2004a)

The architecture to the solution as I propose it is caught in networking architectures. This came forth from the seemingly silly notion of having to build a "nerve system", or basically networking functionality, to control a robot. *Wouldn't it be fun if the network was actually mapped onto our nerve-system?* From this followed that networking models could actually also be applied to our nerve system. Funnily they could also be applied to other types of consciousness, for instance social consciousness. So I took the most general networking model, the OSI RM, and created the CRM from this. This is where it started.

In Sloman's lecture he also describes a *meta-management* system which would then use interrupts, the ignoring of interrupts and other system-architecture related mechanisms to decide which process may continue. Before describing beings as some kind of system architecture in a single computer, we need to realise that these interrupts are *on-off* mechanisms. We don't have buttons which you can mentally push to turn something *on* or *off*. Yes, we can poke our brains and induce the stimulation of muscles in our arms or legs to raise those limbs. But even then the amount of stimulation such a limb receives is dependent on the poke. Muscles aren't turned fully *on* or fully *off*. Just as well thought-mechanisms are neither, because if they were we would never wake up again after having turned them *off*. They may only be put to sleep, i.e. approach an *on* or *off* state. Although we will perceive them to be so, they can never obtain such a level. Just as well if our muscles were stimulated to the fullest degree they would most probably tear themselves apart.

The GToC provides the functionality of these *meta-management* systems which comes forth from the basic architectural constraints of a being. So there's no need to implement it any further. It is already everywhere in the whole of the management system which gives rise to consciousness. Particularly the DS layer and the freeing and blocking of receptors influence or basically provide *meta-management*.

In his lecture Sloman was specifically on what I feel to be a right track and had the same feeling of what should be done. The expression of his notion, the choice of words and the idea of modeling to my opinion however should still be relatively extended. This may be due to a lack of experience in systems design and related subjects. Were he to explore a bit further in systems design, his expression of solutions would have been more agreeable to me. His basic notion, with which I will conclude the review of his lecture with here was correct though.

> *"I expect that the systems approach will show that* ***architecture dominates mechanism****."*

Regarding the parts which I read of Sloman's book[10] I have nothing to mention which hasn't been mentioned for his lecture.

4.5 Rolf Pfeifer and Christian Scheier

Pfeifer and Scheier in their book "Understanding Intelligence"[11] basically describe where we are now. I haven't read much of the book because there was already so much to read. Skimming through it quite some issues were dealt with

[10](Sloman, 1978)
[11](Pfeifer and Scheier, 1999)

like for instance artificial evolution and a framework for a theory of intelligence. I'm going to focus on the last pages of the book, since my thesis begins at the end of quite a lot of research which has already been done.

The theoretical framework for a theory of intelligence as proposed by Pfeifer and Scheier contains a few aspects. External aspects of the framework are a theory of intelligence, technology applications and implications for society. These come forth from a frame-of-reference, time perspectives, diversity/compliance and design principles. At the core we find desired behaviours, the relation between the agent and the ecological niche. This should then all be captured in a synthetic methodology. Can this be related to the GToC?

The theory of intelligence is caught in the GToC by establishing associations which may grow more and more complex. Technology applications don't necessarily have to be stated in a GToC because understanding ourselves is a good enough goal itself. Seen from my background in information technology however I have illustrated a few examples of derivative technological possibilities. The implications for society are on the one hand the better understanding of society. On the other what's now needed and can be found in *appendix I* is a generalisation of ethical boundaries for all beings.

The frame-of-reference is constituted by our physical knowledge of beings and the world around us. Just as well to a limited extent a physical interpretation of self-reference can sometimes partly be used to identify certain aspects of being. The time perspectives come forth from social structures and changes in relations within a being's society or habitat to which it has to adapt. The diversity/compliance comes forth from evolution and the design principles are caught in physical laws.

Desired behaviours come forth from a being fitting into a specific task trying to maintain itself and its species. The relation between the agent and the ecological niche is formed by the being's place in the PW.

The synthetic methodology has been illustrated by the statements and definitions as well as the graphical models. The way the model should be handled is illustrated in the rest of the thesis.

One question still comes to mind: does intelligence mean consciousness? This highly depends on the values which are attributed to the word *intelligence*. Pfeifer and Scheier acknowledge this by stating that intelligence cannot just be measured by for instance IQ-tests. It largely comes forth from interaction with the world. What consciousness means to me is the whole of the interaction with the world, intelligence and acknowledging being part of the world itself as caught in a single being. This for a large part can be treated by philosophical debate, but doesn't undermine the principles of the Consciousness Reference Model (CRM).[12] This model is physically correct and can be used to explain conscious behaviour.

4.6 Owen Flanagan

Flanagan in his paper[13] proposed to use a natural method of analysis to come to a theory of consciousness. The way he describes his natural method is collecting all data and then searching for coherence between this data. Most unfortunately

[12] Chapter 5

[13] (Flanagan, 1997)

this doesn't say anything about how the data is examined. To make it even worse, life is all about just learning from what we hear and see. Stating that we're using his natural method doesn't say anything else other than living our lifes and keeping an eye open for consciousness.

The *application of the natural method* fails technically speaking by its own definition. It states that all ideas from all research areas are collected. But by applying *the natural method* to a specific problem it narrows down the research field even though quite often the solution can be found in a seemingly unrelated research field. That *application of the natural method* fails is quite clearly illustrated by the size of Flanagan's paper. Had he gathered all information from all views, his paper would never have been restricted to such a limited amount of pages. He quite clearly made a choice of papers which he felt were relevant and which weren't. But this isn't reflected in his paper.

In case of research people should understand that research depends on ideas and not methodology. Although someone may have a perfectly sound idea, he doesn't need any methodology to have that idea. He will have to search for a method to display his idea, but this method depends on the idea itself and is thus variable. Suppose that we would choose a method before having the idea, the idea would have to be fitted into this method. But not all ideas fit into the same method. What I would like to state is the following:

statement 4.1 (method of research) *A method of research doesn't lead to ideas but comes forth from the notion how a specific problem should be solved.*

This of course means that the method chosen for a specific idea is only a valid methodology if it directly leads to the answer. When it indirectly leads to an answer it was part of either a larger process or a larger method.

I don't accept theories, because these would be limited to single existence-instances. Single existence-instances should not be categorised as theories generally applicable in fields, they should be seen as examples. Examples cannot prove a general theory they can only acknowledge a single instance of an observed fact. In case of general principles in accordance with Popper[14] I state that there is no proof.

Flanagan's methodology fails because it doesn't propose a framework which can be used to describe subsequent processes of consciousness. It also fails because it just states that we should listen to ideas from all fields, but it fails to identify those ideas. It even fails to identify in which part of different research areas those ideas supposedly should be found.[15]

The main problem is that most people haven't got a clue what consciousness is about and don't have a single coherent clear thought about consciousness. Having such a thought would reduce us to very simple mechanisms and quite possibly take away the *superiority of mankind* which quite a lot of people think we have. To most people it's just too damned scary to even have a concrete idea about consciousness.

Flanagan also speaks of different kinds of consciousness which are again specialisations of consciousness. But shouldn't we first look at what consciousness

[14](Popper, 2004)

[15]His answer should then be that we should look everywhere. By only listening without active search, without often even knowing what you're searching for, it's hardly possible to gain enough ideas during a life-time to have an original idea yourself.

generally contains before we can specialise? The natural method did mean searching for similarities, but things like evolution and different types of behaviour quite easily distracted Flanagan from the more general goal. This shouldn't have happened. First there needs to be a theory then we can look at the consequences.

4.7 Patricia Smith Churchland

Churchland in her paper[16] chooses to examine the nervous systems as present in organisms. She chooses a self-defined reductionist view, which is coherent and thus clear in its definition. The reductionist view is illustrated by a hierarchy containing seven levels of abstraction. From this hierarchy alone quite a lot of the functionality in the CRM can be derived. Even the basic idea of the networked architecture of the CRM is represented by the hierarchy.

Churchland describes both consciousness and motor control as being problems in the building of an agent. Unfortunately she doesn't bind them in an agent architecture. I share Pfeifer and Scheier's view[17] that consciousness cannot exist without a properly motor controlled morphology and vice-versa. This morphology should be properly arranged to be able to explore the world.

In her search for consciousness she describes specific fields and multiple hypothesis which can be correlated. Different mechanisms are identified which should be studied. The only thing I have to object to this is that it would probably lead to theories, but they would be largely non-implementable. By working the way she does it would in the end have led to the most basic of mechanisms which should be identified in describing consciousness. Eventually this would have led to the proper theories, so I have to admit that she would eventually have been correct. Although I feel that she could have taken a shorter route, maybe people like her put me on my route as well and thus I follow on her route. I'm not sure and don't think I can be.

[16](Churchland, 1997)
[17](Pfeifer and Scheier, 1999)

Chapter 5

The General Theory of Consciousness

In chapter 2 I've introduced the basic outline of the CRM which should eventually lead to a conscious being. The layers already give a broad view of what kind of influence the world and the cognition of a certain being have on each other. Still, without loosing the abstract idea into which different models of consciousness should be fitted, it is possible to go down even further in a more detailed explanation of each of the layers. This will mean describing the different interaction mechanisms of these layers.

Although the processes can theoretically not only start running from the PW layer, but also from the CO layer, each of the layers is built on top of the other. Here I mean that the PW is the layer on which all the other layers functionality run. So in order to describe the layers fully it's best to start from the PW layer and work all the way up to the CO layer. Just as well the CO layer may induce a process back down to the PW layer to influence the PW. Having seen these processes it's possible to make a bit more of a distinction between the mental processes and the physical processes. The distinction can be discussed together with some ideas of how to fit all this into different research areas.

It's important to note that the text is often just meant to illustrate how I came to the keywords, assumptions, definitions and statements in this chapter. The text often contains a large amount of ambiguity regarding the terms used. The keywords, assumptions, definitions and statements shouldn't contain any ambiguity. Neither should the graphical models.

5.1 The Possibility of Perception

As we all know from experience, signals are emitted by the world to us which we perceive. But how should we fit this into any theory of consciousness? Not all the emitted signals are perceived by us or any kind of being in general. Some are perceived by neighbouring beings, some not at all. So basically a signal has the possibility of hitting a being.

All the signals that haven't hit anything at a specific moment in time have the same wandering property. When a signal is wandering there's no way someone can predict for sure that the signal will hit something else. The only thing that

can be stated basically is what has already happened. Hereby we assume the following.

assumption 5.1 (knowledge) *It isn't possible to know everything that happens in a world, so you can never be sure which signal is going to hit something in the far or near future and which not.*

It's of course always possible to make a reasonable guess, but even if the chances that something will happen are approximately zero, it still may happen. And approximately zero is as small as a chance may get. Sometimes people define chances to be zero, but even this definition is just a guess based on highest likelihood to what it may be.

5.2 A Generalisation of the Applicability of the Model

Until now the text has been mainly concerned with beings as a single CO structure. The model should however be applicable to any size of being or CO structure.

What I here now state is that the model should be generalised in it's application. So it should not only be limited to beings which are a single CO structure, but it should also describe the smaller CO structures that make up a single being. This can for instance be neurons. Just as well multiple beings can decide to solve a problem together and then form a new CO structure together. The model can also be applied to behaviours within society, where all the participants in society make up the larger CO structure of society.

No matter how large the number of smaller processes making up a CO structure, the rest of the text will refer to these structures as *a being*. So a being may be as small as a neuron, but a being may also be considered as big as society.

5.3 Receiving Signals

What we basically know now is that a signal may or may not hit a being. This means that we'll have to make a distinction between signals which have and haven't hit a being at a certain point in time. What we now should have is a certain signal where it may be sleeping (or dormant to use other terminology) or it may be in the state of hitting something, activating it. To clearly make a distinction between these two states, I hereby introduce the following two definitions.

definition 5.2 (dormant) *A dormant has the possibility to in the future hit a receptive and raise that receptive's charge.*

definition 5.3 (activator) *A dormant becomes an activator the moment it hits the receptive and raises the receptive's charge.*

As you may have noticed a receptive can be anything, even a rock. Suppose a light particle hits the rock, actively raising its temperature. This doesn't necessarily have to have an effect on a being. On the other hand the light particle may hit for instance the being's retina or perhaps its skin. Now it does

play a direct role in perception. We are only concerned with influences directly concerned with perception.

Before a dormant may hit a being; the dormant needs to be emitted into the world. The single fact that the world may contain dormants is illustrated in figure 5.1. The set of dormants contained within the PW layer is passed on from one moment to the next.

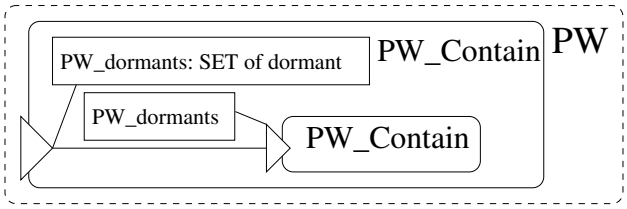

Figure 5.1: The PW contains dormants

Just as well the PW contains receptives. Or to be more precise it contains physical structures on which certain relations hold that we here call beings. These beings have receptors which are instantiations of receptives associated with perception. A formal definition of a receptor is given in definition 5.4.

definition 5.4 (receptor) *A receptor is a receptive that constitutes part of the physical relation that is a single being.*

Figure 5.2 illustrates that a being has receptors. Each receptor has a specific charge associated with it as well as a maximum charge it may contain.

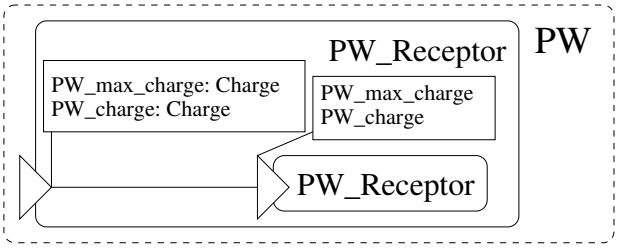

Figure 5.2: The PW contains receptors

As can be seen in figure 5.3 the dormants are emitted into the PW layer. Before a dormant may become an activator it first needs to be emitted. After a dormant has been emitted possibly another dormant is emitted. Just as well any dormant may become an activator. What should be noted is that it isn't certain that either one of the two will happen. It's possible that a dormant is emitted or that a dormant becomes an activator, but it isn't certain whether it actually will. As can be seen, the emitting of new dormants and the becoming of activators of previous released dormants are processes that run in parallel. Regarding the influence on a being it's important to now realise what's stated in statement 5.5 about the activation by former dormants.

statement 5.5 (perceiving through receptors) *In order for a being to perceive an activator should be in a direct relation with a receptor as found on that particular being.*

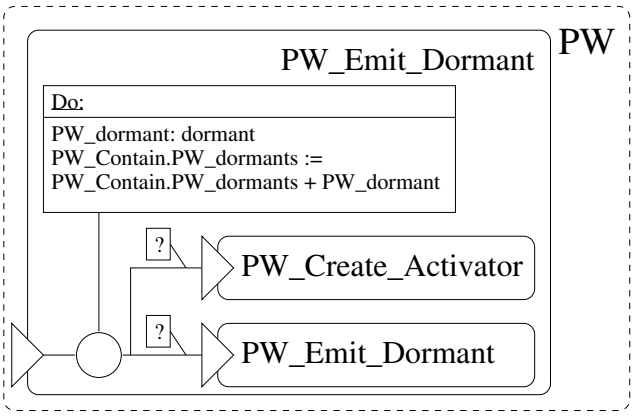

Figure 5.3: A dormant is emitted into the PW layer

In describing how a dormant hits a receptive and thus becomes an activator, I mention the word *charge*. Now what does this mean? There are different ideas one might have regarding the word charge in relation with cognition. When we are stimulated more often per limited time-frame the experience becomes heavier. Is this what I mean by charge? No. What I mean by charge is that for a receptive to emit anything itself, it should first be properly stimulated. The heavier experience thus comes forth from the receptive firing signals more often per time-frame after being properly stimulated. But this is of later concern.

A good way to picture how a certain receptor might work is drawing a parallel with a capacitor. This first needs to build up a charge before it emits and if it isn't built up fast enough the built up charge may leak away.

statement 5.6 (receptive charge and emitting) *For a receptive itself to emit, by being hit often enough the receptive's charge should first be raised to rise up to the maximum containable charge before that receptive can emit.*

Figure 5.4 clearly illustrates statement 5.6. A dormant is turned into an activator the moment it hits a receptor. Although any receptive is contained in the above theory, it's not always directly associated with a being. For the description of the physics of an environment the science of physics should suffice. So here we just concentrate on the specific instance of a receptor. As can be seen the happening of the transmission depends on the presence of a charge that exceeds the maximum containable charge. Only then it may emit in order to transmit the signal. But what does it emit?

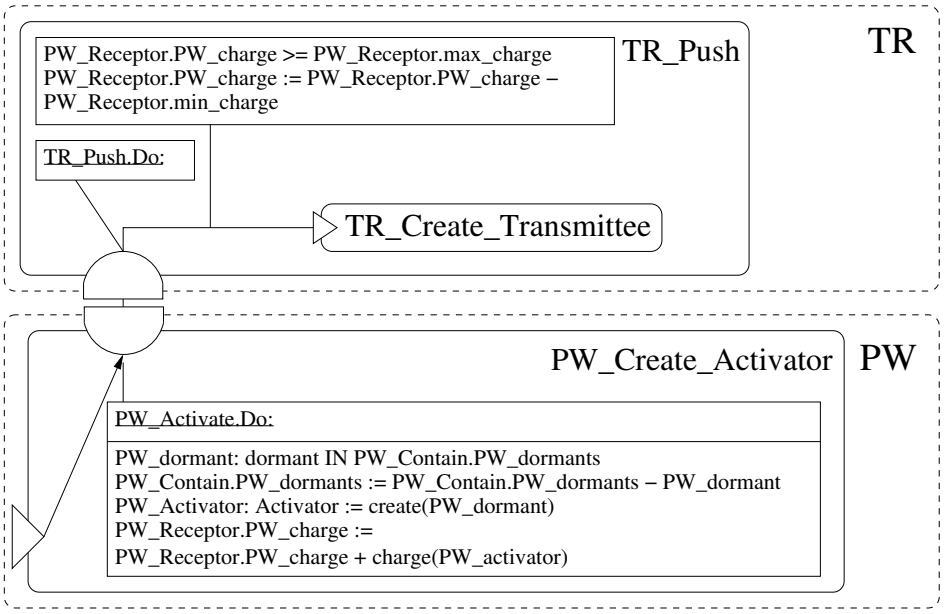

Figure 5.4: A dormant hits a receptor, becomes an activator and raises the receptor-charge

5.4 Propagating Signals over the Network

In the sense of the world, light may hit something, creating warmth. This warmth is then just emitted into the world again and doesn't have any relation to any kind of information. A being may establish that the light created the warmth but this follows from causality theory, not from containment of information within the warmth.

When light hits an organism, the organism may fire a different kind of signal which may be classified because it's structured in some way and is meant for information propagation. Such a signal is usually made up of a multitude of smaller signals. Because there exists a relation between the smaller signals that make up the information content of the larger signal I hereby propose the following definition.

definition 5.7 (transmittee) *A transmittee is an encoding of certain information passed by dormants between different processes and has been emitted by a receptor onto the network.*

The transmittee as defined above is sent over the NE. This can be seen in figure 5.5. As can be seen a transmittee is emitted onto and thus transmitted over the NE. The way the transmittee contains the information now of course differs for each type of NE. So a transmittee has to be formed using a common language applicable to the parts of the NE onto which it has been transmitted.

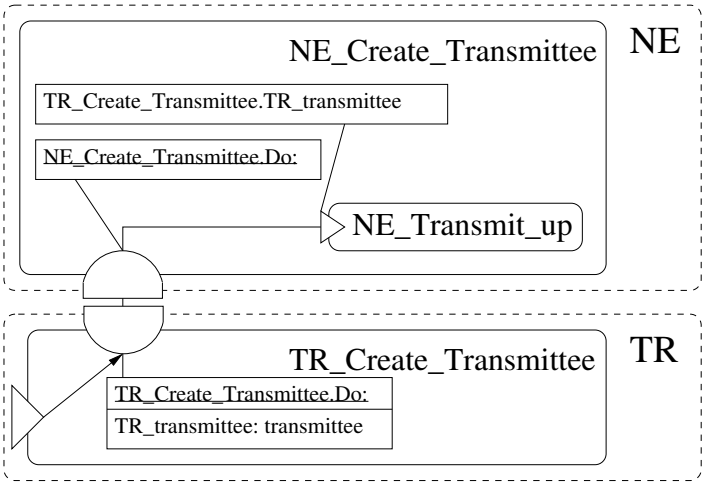

Figure 5.5: A transmittee is transmitted onto the NE

5.5 The Travel of Signals

When a signal is passed over the NE it's important to know where and when it originated from or should arrive. The information always goes to the CO process when it comes from the PW. Just as well it comes from the CO process when it goes back to a certain part of the PW. Because of this only one specific space-time location needs to be accounted for. Or to put it differently: I do care about my senses, but I don't actively perceive anything of them being processed in my brain. I don't perceive the functioning of my brain and thus not the time and space where signals arrive in my brain.

The NE itself is basically a collection of processes that are directly or indirectly connected. Because of the differences in implementation of these processes they may each have their own language of passing information. It can be compared to the differences between for instance analog and digital signals. As such the NE also defines how the information is passed, i.e. the encoding of the transmittee.[1]

Because there may be many routes to get from one point to another the information should be passed using the shortest possible route. In case of human beings this is just the way we are built. In terms of society news spreads according to certain diffusion models. Because of this the network will find the shortest route from one point to another because it takes all routes. This route just happens to deliver the news the soonest.

definition 5.8 (network) *A network is a set of directly or indirectly connected processes using an agreed upon kind of information encoding to send the information over the shortest possible route.*

When the transmittee is sent over the NE it should then be well established what the time and location of a certain event were or should be. In on the one hand perceiving relations and on the other executing actions it's very important

[1] Section 5.4.

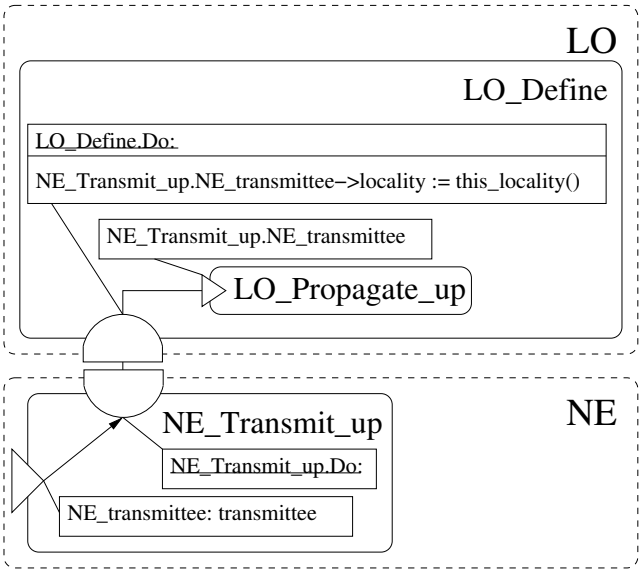

Figure 5.6: A transmittee's LO is defined by the LO of transmission

that the place of an event as well as the time is clearly defined. This is necessary to determine the order in which things happen. Just as well it's important for the coordination of tasks to execute the right actions at the right moment. This is illustrated by definition 5.9 and figure 5.6.

definition 5.9 (locality) *The locality of an event is defined by the space-time coordinates of that event.*

5.6 Collecting All Signals

When it's clear what the LO is, it's possible to make a selection of the incoming signals to which a being would like to listen. Or more precisely per signal it should be well established to what extent a being would like to listen to that signal. Of course sometimes in order to throw a signal away it should even be perceived the smallest bit thinkable. So it's not always possible to fully block any signal, although the amount to which some signals are listened to may prove to be nihil. Of course a signal is handled in unison with the rest of the signals in its time-frame. The gathering of signals is shown in figure 5.7.

The transmittees sent from the PW layer are contained in the DS layer after reception from the NE. The DS layer is also where they are processed. The fact that they are contained in the DS layer is modeled in figure 5.8.

For abstract problem solving a being should be able to filter out the right pieces of information and leave the rest. As stated before, the more often a signal is emitted the more intense (or heavier) the experience gets. If receptors over a broad range receive too many stimuli it may not even be possible to make a clear selection of the incoming signals. When we for instance look at two people having a conversation, the one person concentrates on the other person's words. When they are in an empty sound-proof room, this doesn't pose a problem.

When they are in for instance a disco with a lot of noise, then it becomes hard or maybe even impossible to hear the other person. This happens because too many stimuli reach these persons ears. Just as well when they are in a room with a lot of different people speaking, it may also be impossible to hear the other person. This was more generally called *the Cocktail Party Effect* by Handel.[2] But how do we model these kind of effects into the DS layer?

The best way to model the perception of signals is to again attribute a certain potential to all the signals. Here the potential then represents the perceived heaviness of a signal within a certain time-frame. The signal that induces a potential peak relatively high to the other signal potentials will be the one that is heard the loudest.

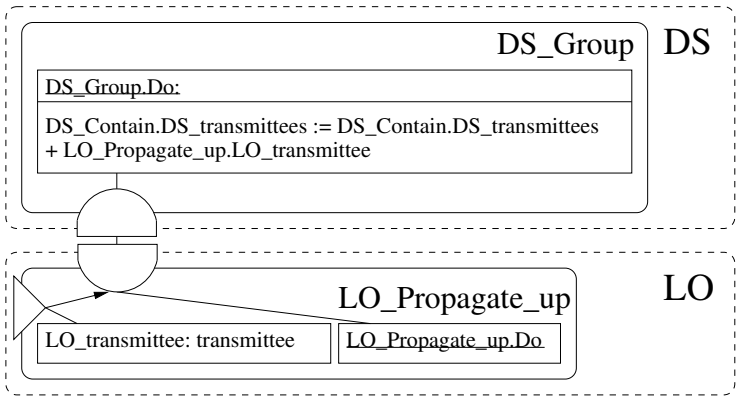

Figure 5.7: A transmittee is added to the set of transmittees so it can be combined into one signal

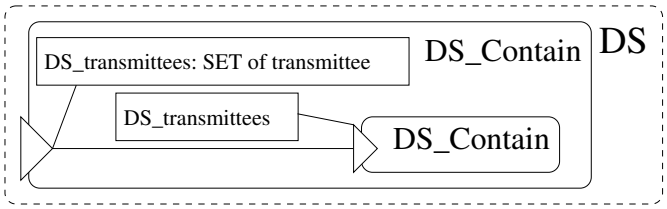

Figure 5.8: The DS layer contains transmittees

5.7 Information Representation

Eventually the CO layer will receive a certain range of potentials as its input signals. Each potential is a function of the heaviness of the incoming information as contained in the signal as well as the aim of CO focus or *desire*. If you try to concentrate very hard, focusing on a certain signal, then the chances that you will perceive it are relatively larger. As stated in section 5.6 the heaviness can be crudely derived by looking at how often transmittees containing a certain piece of

[2](Handel, 1989) and (Hobo, 2004b, Section 5.1)

information arrived during a certain time-frame. The information instance that is generated from this contains the heaviness of the information as a property.

definition 5.10 (heaviness of information) *The heaviness of information is the frequency with which a receptor emits transmittees containing that information.*

The gathering of information and focusing on incoming signals is a continuous process. This is shown in figure 5.9. The gathering and focusing means the following. When information is gathered, immediately a certain focus in the form of the potential wave is built. This is a process induced by the CO processes as well as the PW. The CO process utters a desire to listen to certain signals and the PW may or may not crudely disturb these desires. When the PW emits a signal that's really heavy this may mean that a being is forced to listen to it. Whether a beings is able to ignore the signal depends for a large part on strength of will. Strength of will is now reduced to a physical property to build up resistance to disturbing factors. More on this in section 5.14

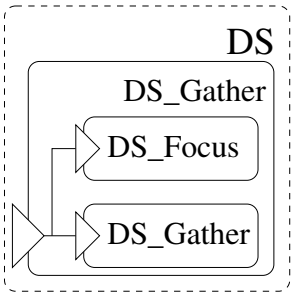

Figure 5.9: Consciousness needs a continuous gathering of and focussing on the transmittees

Each focus determines a certain time-frame with a certain length. The multiple foci determine the perception of the PW around us, where this process of perception expands over time. When a time-frame has a finite length, this also has the implication that as the time-frame shifts, not all transmittees are instantly forgotten. Time-frames may then overlap. Of course this is just a crude representation of the PW.

In reality a time-frame doesn't shift, but the signals fade away. There's no such thing as the perfect switch in turning *on* or *off* a lamp. The lamp always goes through a warming up and cooling down process, even if it's not noticeable to us. This of course means that in such a model all previous signals are remembered to a certain extent. Even though their influence will approximate zero when time goes to infinity, it will never truly become zero. I cannot model options like this into the CRM. These are just suggestions for instantiations of the CRM.

So what we now have is a collection of potentials corresponding to different signals. This will be part of the final signal, where the final signal will be formed partly influenced by the CO desires.

The focus of a collection of potentials can be determined by looking at the relative height of the potentials. The potential peak of a signal has to rise relatively high above the other peaks to be perceivable. Of course the focus depends on both the desire of the CO process and the PW influence. The final signal can be determined by looking at each single potential instance and the corresponding desire. Based on these two properties the final signal in the form of potentials may be derived.

definition 5.11 (information potential) *The information potential is a function of the heaviness of information and the desire to focus on that information.*

definition 5.12 (potential wave) *The potential wave is a continuous approximation-function of the collection of all information potentials according to their contained information.*

This of course means that in generating such a potential wave there may be a clear interaction with the CO layer. However, this need not be true. Suppose that a signal is unexpected? If a signal is unexpected, it will automatically gain a very high potential. This will largely be due to the fact that the resistance for such signals is very low. In selecting to listen for a certain signal, a being expresses the desire to listen for a certain signal and builds up a resistance for signals it doesn't want to listen to. But suppose a new signal arrives and no resistance has been previously built up? Then this signal will have a very high potential disturbing the previous focus. If you don't know that a certain signal will exist in the future[3] and needs to be ignored you will not build up any resistance for that particular signal.

Only when actively not associating with any signal, a being isn't aroused by an *unexpected signal*. In such a case a being isn't expecting nor not expecting any signal or, to put it differently, expecting not to expect. This means that the signal isn't unexpected at all, because for a signal to be unexpected there must be an association of what to expect.

The physical representation of non-association would be the free flow of signals where the association processes are halted. So the resistance for the actual association processes then is really high, but the resistance for the experiencing processes in terms of feeling is very low. As you may now see this process of non-association contains multiple smaller processes, where the process of non-association itself will look to the outside world as if it has a free flow of signals. This is because the flow is the only perceivable thing. This is just an example of a small problem regarding explanations considering the complexity of different instantiations of the CRM working together.

statement 5.13 (unexpectedness of a transmittee) *Unexpectedness of a transmittee claims focus on that transmittee, where not expecting comes forth from expecting the complement of certain received transmittees.*

statement 5.14 (focusing on a signal) *Actively focusing on one signal means actively diminishing the focus on the other signals.*

This last definition permits the unexpectedness and expecting not to expect to happen by creating relative foci on signals. In unexpectedness the focus for the unexpected signal becomes relatively high on receiving the signal, because the PW influence is very high. The receptors don't have any previously built up resistance to the signals which is higher than expected. Because of the high flow of signals this will produce a certain potential. This in combination with the associated desire which doesn't ignore the signal will easily produce a clearly perceivable and influential potential.

One thing you should note before proceeding is that the actually generated potential wave contains all the information that's processed by the mental processes. It contains the location where a signal originated, it is perceived in a certain time-frame and most importantly it defines the conjunction between

[3] Assertion 5.1

focus and heaviness, i.e. achieved focus. So the mere generation of such a wave-function captures the whole of reception of signals and the focus in one single instance in the mental processes of a being. How the potential wave is generated is shown in figure 5.10.

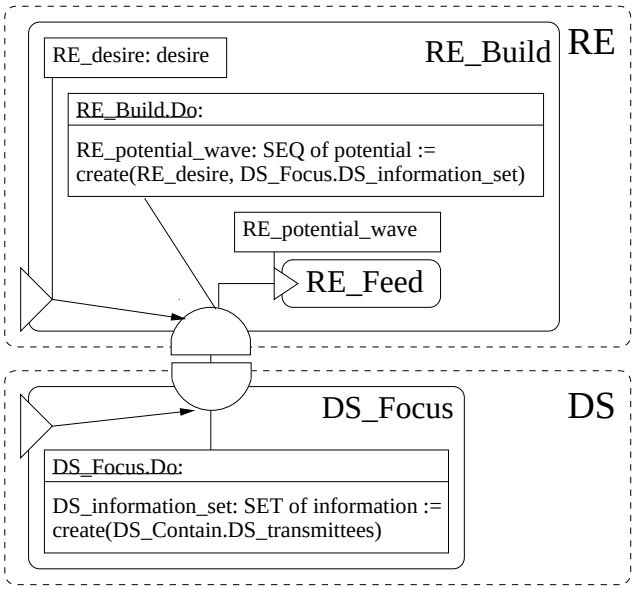

Figure 5.10: The potential wave is built

First all information is gathered by merging all signals with similar localities defining the heaviness of information and attributing it to this information. Suppose a human being would rub its finger. One signal from its finger will not be enough to establish that it is rubbed. On the one hand the being needs to establish that a larger area is stimulated. On the other hand it needs to establish that it is stimulated over a larger period of time. The haptic senses associated with rubbing now have initiated this process of the sensation of rubbing where in a certain time-frame the stimuli are collected and each new time-frame remembers that the previous time-frame also contained the same kind of stimuli. The memory comes forth from the recurrence of desires.

Second, based on the gathered information as well as the desire to focus on a certain signal a potential wave is generated. The potential wave now becomes the representation of the information that the being is fed by the PW. The potential wave thus contains current data and previous data caught in the desires as well as desires to initiate certain changes in or continuations of relations.

You could try to envision the potential wave as shown in figure 5.11. As you can see it could contain a minimum potential which has to be achieved and it could be divided into five areas representing each of the senses. In certain areas the overall potential will be higher than in other areas. This way by assuming the same potential for certain types of information, the correlation between different types of information may be recognised by the CO processes. Remember that this is just an example picture and not something that has been or can be measured. We do know that some kind of potential exists. It is

possible to stimulate certain parts of the brain to induce a potential, sometimes even just by poking it.

Since the potential wave is partly a representation of the signals fed to us by the PW, another way of looking at it would be definition 5.15.

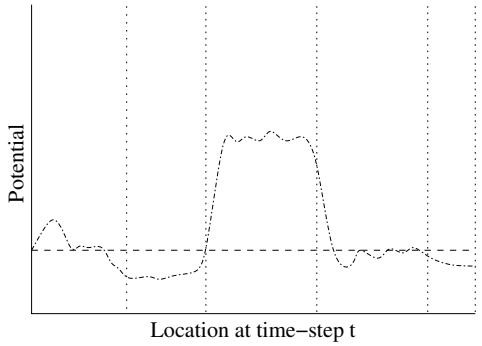

Figure 5.11: An abstract representation of a possible potential wave.

definition 5.15 (representation) *A representation is a unification of multiple transmittees as well as desires into one new specialisation of a transmittee that suits the input requirements for the cognitive layer.*

5.8 Desiring to Stimulate

The CO layer receives representations in sequence. Each of these representations may then induce a step in the process of cognition. But what is this process of cognition made up of? Manzotti[4] proposes a mechanism which should lead to consciousness which for a large part also contains cognition. It's a so-called process based architecture for an artificial conscious being. He describes a Basic Intentional-Robotics Unit (BIRU). Since I propose a more generalised architecture that should be applicable to any conscious being, I also need to generalise the mechanisms as described by Manzotti. Manzotti identifies three basic learning mechanisms which need to be seen from a unified point of view for a single learning structure which he uses for his BIRU.[5] The three learning mechanisms are based on the following three assumptions that need to hold.

1. **Relative similarity.** This means that for a new unit of information there may not have been a previously learned representation for it already. Something new thus is something not actively seen before. This may of course have a relatively high unexpectedness associated with it, but when exploring actively, this need not be so.

2. **A priori learning curve.** This means that learning may take place during certain stages of a beings life. During these stages information is basically imprinted on its mind. Later in its life it will then be slower at or not capable of learning.

[4](Manzotti, 2003b,a)

[5]I explicitly state learning structure and not neural network. A neural network is already an instantiation so I'm avoiding that term here.

Different theories could be formed regarding this point. One of them would be that the signals which induce the stimulation of a being aren't passed through by the DS layer anymore, because the desire is too low. This stimulation is discussed as the next item. The paths of connections between the CO processes through multiple instantiation of the CRM could also loose their flexibility. Or maybe the system runs out of memory. These are all things that can't be modeled explicitly, because it's a specific instantiation of an architecture. The a priori learning curve, or the ability to learn, thus has to follow from the underlying architectures.

3. **Significant stimuli.** Significant stimuli are uttered during the reception of a certain signal, where these stimuli may or may not be pleasing. In more familiar terms of computer science this represents a sort of reinforcement learning. The difference is that it isn't predefined by the designer's assumptions regarding what is pleasing and what isn't. This needs to emerge from the learning being's structure.[6]

Although in all three assumptions learning is involved, it's important to realise that only with the first two it's a pure internal process. Stimuli however, are presented in some way by the PW. These stimuli may be directly presented by the PW layer. They may also be instantiated by actuators[7] where the CO process itself induces these stimuli into the PW layer. Either way the stimuli propagate up through all the layers starting from the PW layer and ending at the CO layer. When the first two mechanisms grant learning, the third induces either a learning process or maybe a confirmation of a previously found notion. When a previously found notion is again affirmed, this belief may of course grow stronger.

But how does a being learn without direct stimuli from the PW? To be able to do this, learning should be pleasurable by itself. This doesn't mean that a being should learn the same thing over and over again just for the fun of it. What I propose is that when a being encounters something unexpected, this should induce an exhilarating effect. When a being receives a lot of signals within a certain time-frame, this will also be beyond normal experience and thus at least physically unexpected. A parallel could be drawn with adrenaline, the faster beating of our hearts and the heavier breathing and thus relative higher amounts of oxygen. This unexpected thing may then be an encounter with a strange notion or idea or it may be a more physical event. This way the being is always provided with stimuli which have no true predefined notion on what's happening. It resembles the natural mechanism as defined by evolution. Another example of the pleasure of learning would be the recognition of a joke. In the beginning it's fun, but after hearing it too often it may become boring or even annoying.

definition 5.16 (newness) *The newness is a function that's reversely proportional to time.*

statement 5.17 (conditions of learning) *Learning can only arise when the following three conditions are met:*[8]

[6] (Hobo, 2004b, Section 3.1)
[7] Definition 5.25
[8] (Manzotti, 2003a)

1. *The being is not familiar with the encountered, i.e. the encountered has a high newness to it.*

2. *The being is in a phase of learning.*

3. *The being receives certain associative stimuli.*

definition 5.18 (unexpectedness) *Unexpectedness is the happening of an encounter without the previous notion of that encounter's happening in the nearby future.*

statement 5.19 (stimulus) *A stimulus is either provided by the physical world or induced by unexpectedness.*

Stimuli which come forth from the CO process need to be induced into the PW. A stimulus is a kind of dormant which may become an activator. But until now we've only discussed how signals are propagated up to the CO layer. So what is now important to realise is how these signals propagate back down.

First of all the CO layer produces a certain output. This output is the result of signals previously fed to the CO layer. The output is meant to induce actuators. Although it is meant to do so, the desire to do so may not always be given the priority by the DS layer. The CO process can thus not decide what happens or what is actually happening. The CO process may only express a certain desire where the desire should be high enough to actually be able to induce a certain actuator to be released.

Here the actuator may also be the initiator of a thought like "Eureka!". A thought that can be perceived, just as anything else that can be perceived, is part of the PW. Remember that it's just terminology. A thought may not be physical, but it constitutes a sequence of relations propagated through the PW layer.

Figure 5.12 which shows how a certain input induces a certain desire. This desire is added to the collection of desires in the CO layer.

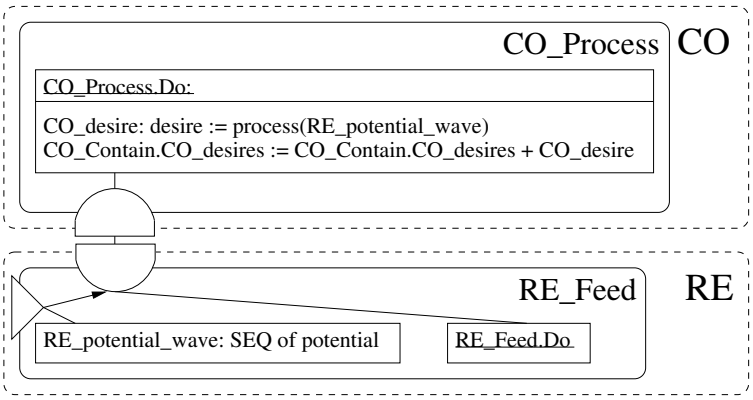

Figure 5.12: The CO process

After the desire has been formed by the CO processes it will try to claim focus on a certain signal as well as induce a stimulation of this signal by certain

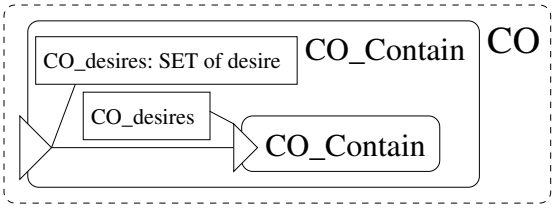

Figure 5.13: The CO layer contains desires

actuators. This will happen in competition with other desires. All these desires will be collected in the CO layer. This is shown in figure 5.13.

Something that should be noted is the way the embedding of associations occurs. Statement 5.17 has a lot to do with this. When a process is repeated often and is less new, the being will not learn anything from it. This way the associative stimuli directly generate a certain output which is hard to overrule with something new since it's built firmly into the CO process. The embedding of associations thus is part of the CO process just as it's part of our brain.

I'd like to illustrate this in case of our brain, noting that this is an instantiation and thus only serves as an example. Our brain consists of parts associated with specific functions. This also means that each part has particular kinds of signals that it's more receptive to than other kinds of signals. Or, to put it properly, each part has a higher desire to perceive certain signals than for others. These parts often learn slower. It's hard to learn to dance. It's a process that takes time. But when you do have it and don't have to think about it anymore, you'll most probably never loose it again. Of course sometimes it's still a matter of discussion whether slowness of learning comes forth from the complexity of the thing learned or from the rigidity of the learning structure. But it's well-known the last part does play a large role. Let's make an assumption.

assumption 5.20 (embedded processes) *Forgetting old embedded processes means learning new embedded processes.*

statement 5.21 (embedding of processes) *In cognitive processes slow learning and slow forgetting leads to embedding of certain processes because of rigidity of the learning structures.*

Assumption 5.20 doesn't necessarily hold, because it fails for instance in case of Alzheimer's disease. Of course Alzheimer's disease means a deterioration of the entire brain by accelerated death of brain cells. This means all mental processes don't function as they should anymore. More on this can be found at the Alzheimer's Association.[9] We'll of course assume that the structure doesn't deteriorate at an accelerated speed in discussing the CRM.

5.9 Competing Desires

When the desires have been collected they need to compete, where a final desire may be expressed which has won the competition. The competition continually goes on, which is shown in figure 5.14.

[9](The Alzheimer Association, 2004)

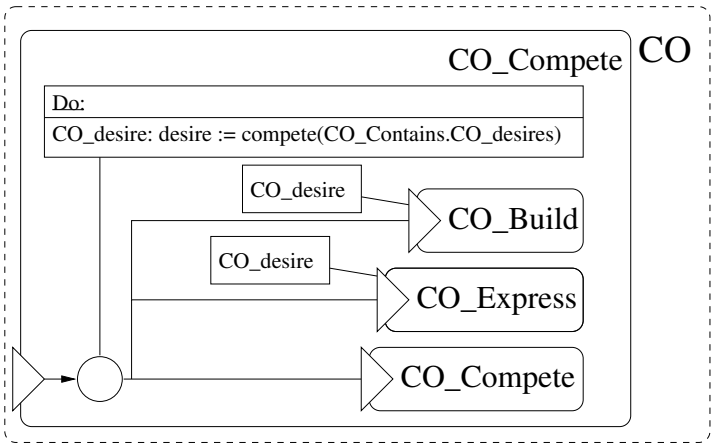

Figure 5.14: The surviving desire is picked by natural selection

The competition to decide which parts of desires may continue to form the final expressed desire is held for every common LO that's defined in each of the desires. This means that the winning desire contains larger or smaller parts of different desires. In our further discussions it will be necessary to state the following.

definition 5.22 (desire) *The output of the cognitive process is a desire which influences the cognitive process and tries to influence the physical world.*

statement 5.23 (competition of desires) *Which parts of different desires are finally taken into account in the decision making process is decided based on competition by natural selection for each of the localities.*

statement 5.24 (desire stimulation) *The resulting desire tries to stimulate:*

1. *the heaviness of information or*
2. *other physical world processes.*

As stated the resulting desire can stimulate the heaviness of information. This means inducing new information which should eventually induce new actuators. These actuators may then influence the receiving mechanisms of signals (receptors). This will then influence the potential wave that's generated from the incoming signals. Note that the desire that's emitted by the CO process is a kind of potential wave itself. It just not includes the conjunction with the heaviness of signals. It expresses the focus you want to express and attain regarding certain signals.

The stimulation of actuators can be induced by expressing the desire of the generation of information. The CO layer should first express its desire to the RE layer. This is shown in figure 5.15.

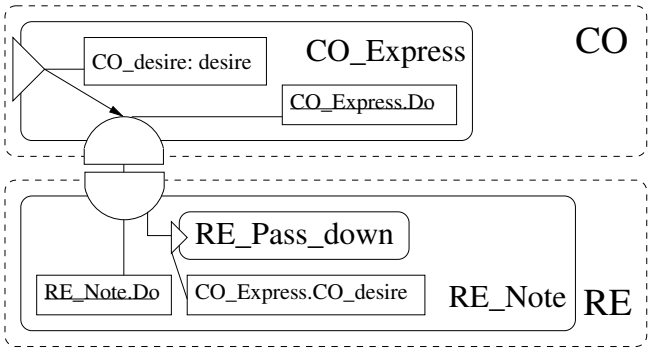

Figure 5.15: Expressing the desire to the RE layer

5.10 Processing the Desire

When a certain desire has been selected, it should be translated into instances of information. This information then induces a further expressing of the desire into the PW. To do this the RE layer may pass the desire on to the DS layer. The DS layer takes the desire and transforms it into a set of instances of information. This is shown in figure 5.16.

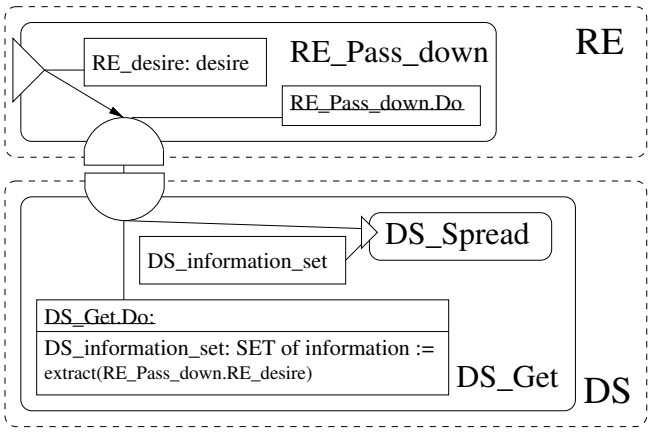

Figure 5.16: Generating information from a passed desire

5.11 Inducing Stimulation of Specific Emitters

When the instances of information have been constructed from the desires by the DS layer, they can be transformed into transmittees. These transmittees receive their LO from the instance of information. This means that a separate process is spawned for each instance of information. Each instance of information is processed in parallel into a set of transmittees. All these transmittees are sent in the current time-frame to make sure that the heaviness of the induced signal is correct. This means that first for each instance of information a process has to be spawned. This process will have to transduce the information as

transmittees. The spawning of the transduction processes or the spreading of information is shown in figure 5.17. It also parses the instances of information to transmittees which can be propagated over the NE.

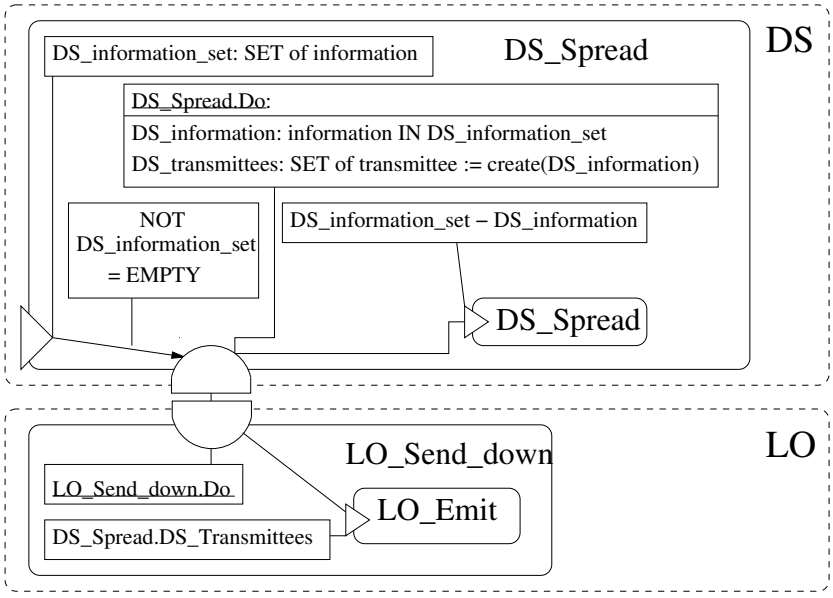

Figure 5.17: The information is spread and transmittees are sent to the right LO

5.12 Sending Transmittees over the Network

When the transmittees have been created from the instances of information, they may be sent over the NE. Each set of transmittees is sent to the specified emitter and has to arrive at the specified time. So the transmittees are sent over the NE to the specified LO. Figure 5.18 shows that they are sent over the NE.

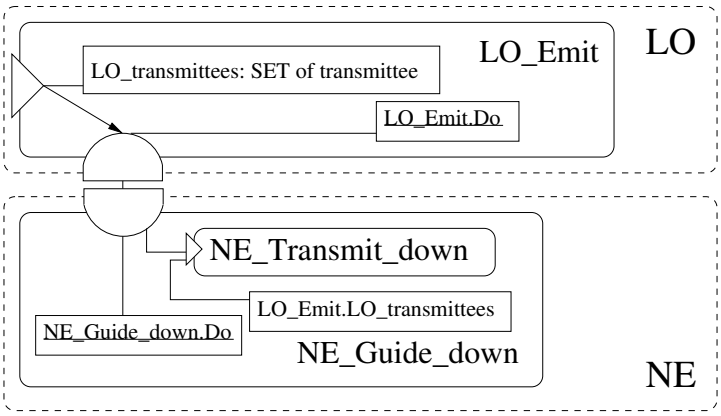

Figure 5.18: A transmittee is transmitted onto the NE which guides it to its TR point

5.13 Stimulating Emission

When the transmittees have been sent over the NE they at some time arrive at the TR layer. This receives the transmittees. The NE doesn't change the transmittees, it just leads them to the right LO regarding emission. This is shown in figure 5.19.

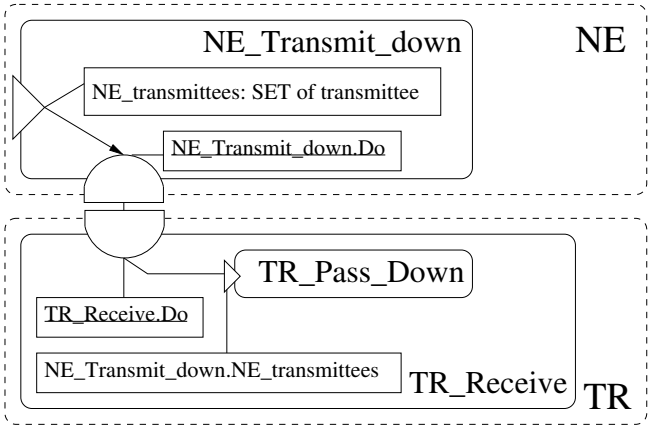

Figure 5.19: A transmittee is received from the NE at a TR point

5.14 Emitting Actuators

Finally the TR layer may process the transmittees, after which an actuator is generated for each transmittee. The actuators are sent into the PW. This is shown in figure 5.20. An actuator is meant to influence the PW. It can on the one hand induce a certain PW process or it may also be perceived itself by the being. The PW process may of course also be the process of receiving signals which is part of a certain being.

definition 5.25 (actuator) *An actuator is an instantiation of a dormant that's emitted by a being.*

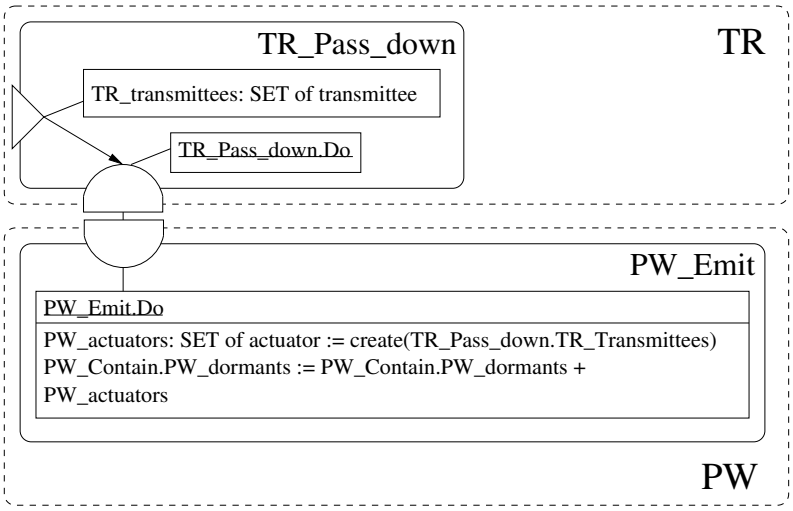

Figure 5.20: The actuators are created from the transmittees after which they are sent into the PW

There are two special kinds of actuators which are directly associated with consciousness. These are either meant for enhancing an incoming signal, i.e. the focus on that signal, or repressing such a signal.

Suppose we are inside a dark room and walk out in the sunlight. Actuators have to be emitted to make sure our pupil closes a bit. If this wouldn't happen, our retina would be hurt by the large amount of light falling onto it. On the other hand it should open again when going back into the dark room.

Another example would be the direct blocking or freeing from receptors which are part of our bodies. This way it's also possible to enhance or repress a signal. This could be used in for instance our brain to enhance the signals spread via certain paths and to repress others. If we want to actively focus on something this is a possible mechanism. However, sometimes we are distracted even though we try to focus on something. This may mean that the signal causing the focus on another particular signal itself was overruled by influences from other signals with a higher signal-potential. This permits us to state the following:

statement 5.26 (freeing and blocking receptors) *In order to enhance or repress the number of activators in a certain time-frame actuators specialised for respectively freeing and blocking the receptor may be emitted.*

statement 5.27 (stimuli and blocking of receptors) *The higher the amount of stimuli the more actuators will be emitted to block a receptor.*

5.15 Explaining Conscious Processes

Something that may be noticed is that the process from going from the PW layer to the CO layer resembles to a large extent that in the opposite direction.

Basically the PW is translated to the CO world and the CO world is translated to the PW. The translation is done by the RE layer. But these signals will still have to compete to be heard based on their potential. This is done in the DS layer.

The DS layer forms a boundary. Here the PW layer can propagate a large influence upwards. A strong desire may propagate its influence downwards. So the DS layer up to and including the CO layer make up the processing units. The DS layer and everything downwards including the PW layer define the connectedness.

On both extremities of the distinction, i.e. the PW layer and the CO layer, the relations are defined in a certain language. From a philosophical point of view it's possible to say that both world images define our relation within the world. Or to put it differently:

statement 5.28 (explaining consciousness) *In explaining consciousness, it doesn't matter if you start at the physical world or the image we have of the physical world, since they are both equal in the explanation that should be given of their representations and thus relational structures.*

The difference with this paper is that the paper is concerned with the processes which may induce consciousness in the end. It isn't a philosophical paper trying to describe consciousness. Since the processes that give rise to consciousness are built on top of or within the PW, this is the start of the discussion held here. But seen from a purely philosophical point of view there isn't really a difference between the two views.

This may have been something that people may have stumbled upon. In describing representations, it can be done physically as in the science of physics, or it may be done mentally, as in the science of psychology. In the former case you have to start from the PW. In the latter it doesn't really matter because they are both equal in representation. This doesn't mean however that either one of the two sciences can't be of an aid to the other.

Just as a matter of convenience I would recommend philosophers and any other kind of scientists to treat consciousness in the same order as done here though. That makes sure that it's easier to see the correlations, since everybody maintains the same order of discussion. That was what the GToC was originally and still is for.

Chapter 6

Emotions

Something associated to a large extent with consciousness is emotions. Consciousness is a basic necessity for emotions to arrive, but not the only necessity. There are quite a lot of theories about emotions. Most identify basic emotional states (BES) and composite emotional states (CES). The BES are then basic emotions which should form all emotions we have. The CES are then containers for different amounts of each of these BES, making up the actual emotion. These models are then used to describe emotions as a classification problem. Is this need to classify justified when you want to discuss emotions? If not, what is justified? What can we discuss?

In this chapter a model is built of emotions, the emotional reference model (ERM), that's in a direct relation with the CRM. It discusses the physical state of a being, the CO influence and the influence of the PW as well as everything that's directly associated with consciousness. Everything that's directly related to consciousness here means the functionality captured in the layers between the CO layer and the PW layer. In other words, the ERM transfers the CRM to the perspective of emotions. By doing so it actually also describes the general behaviour of the system where the emotions are considered to be parts of the output.

The basic relations between the CRM and the ERM are established in section 6.1 by identifying the emotional content. In section 6.2 the way transitions occur between emotions within the ERM is proposed based on the correlations found in section 6.1. From this model follows a discussion on how the representation of emotions comes into being. Section 6.2 deals with the representation or expression of emotions. All these sections define how emotions should be modeled and thus constitute the ERM. Section 6.3 describes the relation with existing models of emotions and modeling emotions as a whole. In conclusion section 6.5 identifies which basic mechanisms need to be there to provide a fully functional emotional mechanism.

6.1 Emotional Content

In describing BES and CES what basically was done in the past is taking a translation of the resulting emotion and trying to reconstruct this emotion from

this translation.[1] What do I mean with a translation?

Emotions are expressed towards the PW. This doesn't always mean that they are perceivable by all beings, but at least they may be perceived by the being which has the emotions itself. To do this a translation of the emotion basically is made into a physically representational state. Then an estimate classification of this state is made. On classifying the state, we start ignoring all the true subtleties and make a crude distinction between different states. We basically remove the fuzziness out of the emotions. This is done even though within a classification there's still a large amount of space left to play with variables influencing the actual emotion. Regardless of this it's mapped onto a stereotype. We then try to reconstruct different emotional states from these stereotypes. It's a bit like translating an English book to Dutch badly and then translating it back badly. It will never become what it once was.

This is unfortunately due to the fact that beings feel the need to classify, to understand, to find basic solutions. In order to come to a decision the input values should be resolved. But the basic solutions are only approximations of the physical solution with a rather large margin for error. The physical solution here is the whole that makes up the emotion which is far complex than any collection of classes of emotions, no matter how extensive. In order to create a believable model this margin will have to become as small as possible. Preferably it will have to be done away with.

We shouldn't look at the result as caught in a single emotion and assume this result to be the beginning of all explanations. Instead we'll have to look at the actual oneness of a being, without any previous influences of other beings or the PW.[2] We need to assume this as the beginning from where the being may develop emotional content based on its own state and the influences of the world. So in order to discuss emotions it isn't desirable nor wise to keep looking at the being in its oneness. This is just the starting point from where relations may develop. The emotions then develop into group behaviour and interaction.

So, where does it start? Before consciousness arrives, before it sets in, what is there that will influence your emotional content in the future? Just before our CO birth there's no CO process, no emotional content that we contain. There's just the world, and the signals it produces which are propagated. Before we look at the CO process after it has set in, we first need to look at the physical properties of the passing of these signals in correlation with the CRM. Our CO birth is basically the moment in time that our brains start working. This may of course be far before wakeful consciousness is obtained.

A signal is passed through the PW. Such a signal in the form of dormants is accepted in a certain amount by receptives. A receptive behaves like a charge. There's a certain flow of dormants that can be established by looking at the number of activations of the receptive. The receptive has a certain resistance in accepting the dormants as activating units. As we well know from physics there exists a relation between the flow $I_{dormants}$ and resistance $R_{receptor}$ creating a certain potential $V_{pressure}$ based on equation 6.1. This potential then basically stands for the PW pressure that's exerted by the dormants. Here $c_{receptor}$ is a constant.

[1] An example of this is the emotional model as proposed by Ortony et al. (1999).

[2] *Oneness of being* here means a being without any previous influence, which has never processed any signal.

$$c_{receptor} = \frac{V_{pressure}}{I_{dormants} \cdot R_{receptor}} \tag{6.1}$$

Now what still is needed is a formal definition of the potential, the flow and the resistance in correlation with the conscious being's physical mechanisms. In case of a conscious being we have dormants, activators, receptors and transmittees. Which of these influence the flow, resistance and potential?

definition 6.1 (receptor flow) *The flow for a receptor is determined by the number of activations for that receptor per time-frame.*

definition 6.2 (receptor resistance) *The resistance of a receptor is reversely proportional to the number of activations per time-frame the receptor can accept.*

definition 6.3 (receptor potential) *The receptor potential is a representative of the pressure exerted by the physical world on the receptor by hitting it with dormants.*

These definitions are basically identifications of processes, where the relation between these processes is caught in equation 6.1. The proof is based on Newton's law that every action is related by an equal and opposite reaction. For instance the resistance of the receptor becomes twice as high. To maintain the same flow of dormants the pressure exerted by the PW on the receptor will have to become twice as high as well.

Here it's easy to make the mistake to say that this potential is the same as the information potential which we encountered in chapter 5. This isn't true. The information potential is a function of the number of transmittees only. This may be highly influenced by the freeing and blocking of the receptor and thus changing it's resistance. But the now generated information potential for a low amount of information with no resistance may be as high as that for high amounts of information blocked by a high resistance. The number of created transmittees then becomes a function of the pressure induced by the PW and the resistance as provided by the receptor. So the information potential doesn't become higher with the resistance becoming higher. It actually becomes lower.

Looking at the CRM you may again notice that there's the physical side and the mental side. The above definitions show how the physical side influences the emotional content. The mental side also influences the emotional content. The mental side consists of the CO desires uttered by the CO processes. Now these CO desires will have to compete with the signals induced by the PW to decide what the emotional content will be. The conclusion is now captured in definition 6.4. The potential wave catches the CO desire and the PW influence in one instance and thus is a suitable representative of emotions.

definition 6.4 (emotional content) *The emotional content is represented by the potential wave where the potential wave will directly induce the physical representation of the emotional content.*

The emotional content directly induces the representation in a physical state of a being by emitting actuators. What this will mean exactly is explained in section 6.3.

What now comes to mind is the following question: is learning an emotion? Although there's a clear correlation between learning and emotions, learning

quite clearly isn't an emotion. An emotion, a *movement of the spirit or soul*, is brought into being by the influence of the PW and the CO desires. It's also the input to learning, where the output is given in CO desires. But the learning is not just the emotion, it's partly also the CO process which runs on the physical structure of a being. What the emotion contains is different stimuli which may lead to learning. (These stimuli will also comprehend for instance the willingness to learn.)

definition 6.5 (learning) *Learning is the process of cognition with emotions as its inputs and cognitive desires as its outputs where the process of cognition is altered based on its inputs and outputs.*

The emotions and CO desires do not imply learning. They may just induce learning and represent what has been learned. But a certain process that doesn't learn (anymore) may theoretically utter the same representation of an emotion as a certain being would. Here the difference lies in the induced change of the CO process. What is an important aspect to note then is that a being learns from its experiences. In our case this also means that we experience the emotions in some way attributing a certain value to the experience. But here we capture the definition of experience in a more procedural view.

definition 6.6 (experience) *The experience gained from emotion comes forth from a change in the structure of the cognitive process from processing that emotion itself.*

Of course this learning comes forth from a certain implementation. One might think of certain processes. So from the point of view of reductionism one might say that even a being is just a complex coherence of different processes and thus a process itself. But this again is an instantiation of the model. How the CO process is defined changes with the point of view taken.

6.2 Emotional Transition

As stated before it's not possible to truly identify the emotional state of happiness. There are many states of happiness and you can't define which state it really is. What will I then describe here?

In this section I will determine the influence of different changes in the PW and the CO process on the emotional content. To do this it needs to be clear what the variables are and how they relate to each other. Because the relations become too complex with a large number of receptors and emitters, we'll just look at a single instance of a receptor and an emitter. This will simplify the relations that we are after, but will still contain all the necessary information. It will reduce the rest of the receptors and emitters to the PW level. When it's clear what the basic relations are it's then possible to look at the derivatives to time. When the derivatives to time are known it's then possible to describe the relations in words as well as what these relations imply.

There are a few variables which are caught in the potential wave and lead to a certain CO desire. In order to understand the way this works more properly the components which are used to derive the potential wave should first be better understood.

First of all the potential wave contains (information) potentials induced by the PW. These potentials themselves are derived from the flow of information, influenced by the resistance to take up information. These potential are basically a function of the PW pressure and the receptor's resistance. Dividing the PW pressure by the resistance gives the flow of dormants. This then is proportional to the information potential.

Second the potential wave contains the CO desire as produced by the CO process. The CO desire itself is made up of potentials, where the induced potential by the CO desires and the PW is then fed back to the CO process. The value of these variables changes constantly over time, where they represent an infinite set of instances of information. When a CO desire is expressed it also induces actuators from emitters. The actuators will then be able to change the resistance and thus the flow as a function of time.

We now know our variables and the existence of their association with time, but how do they change with time? What's the influence of each of the variables on each other?

There are two influences on the variables, namely the CO processes and the PW. The CO processes induce certain potentials contained in CO desires based on previous CO desires and PW potentials.

Before the CO processes set in, it should be noted that the PW influences are the only ones there. As such there are only two variables that have true quantities before the conscious process sets in. These are the potential and the resistance. From this comes flow. Just as well the change in flow, resistance and potential are clearly distinguishable in the PW. From this a few basic formulas follow, which we also know from physics and which may be applied in mathematically modeled system theory. This is all in accordance with equation 6.1.

$$\frac{dV_{pressure}}{dt} = c_{receptor}\Big(\frac{dI_{dormants}}{dt}R_{receptor} + I_{dormants}\frac{dR_{receptor}}{dt}\Big) \tag{6.2}$$

From equation 6.2 you might also derive the change of flow and resistance. This is done in equations 6.3 and 6.4.

$$\frac{dI_{dormants}}{dt} = \Big(\frac{dV_{pressure}}{dt}\frac{1}{c_{receptor}} - I_{dormants}\frac{dR_{receptor}}{dt}\Big)R^{-1}_{receptor} \tag{6.3}$$

$$\frac{dR_{receptor}}{dt} = \Big(\frac{dV_{pressure}}{dt}\frac{1}{c_{receptor}} - R_{receptor}\frac{dI_{dormants}}{dt}\Big)I^{-1}_{dormants} \tag{6.4}$$

All the previous equations in this chapter together with a clock constitute a full mathematical system. This mathematical system may be used to derive basic behaviour of the system. The now known equations only describe the physical properties of the being. So what can be seen from a more procedural view?

What this model of course neglects is the CO influence. It only models the PW influence, neglecting everything that will follow from it. Note that the PW influence is everything inside and outside of a being's body, not just outside. So the structure of a being does have a direct influence on its being itself.

Another thing that should be noted is that this is just a model for one single receptor and emitter. To put it differently: in judging the potential wave, we should only look at the parts of the entire potential wave that are directly associated with this receptor and emitter. This makes things a whole lot easier since we can look at a single instance as proposed.

Suppose that for this instance we have a certain potential which would be the information potential V_i. The CO desire's potential can then be indicated by V_d. The potential as merged into the potential wave is indicated by V_w. The relation between the three potentials is illustrated in equation 6.5, where the symbol $\otimes$ means an associative commutative conjunction of the two signals into one new signal. It's derivative should behave according to standard multiplicative rules of derivation. It can't be based on an additive rule since it isn't possible to obtain negative potentials as output values. A negative potential would lead to a reverse flow, which would turn the outputs into inputs and vice-versa.[3]

$$V_w = V_i \otimes V_d \tag{6.5}$$

Although the real system will be much more complex, the model as proposed here will just model a single receptor-emitter system which will be in a recurrent relation with the PW. This also means that the emitted actuators may be perceived by the receptor (but are not necessarily). These relations are there in our PW but they just constitute a small subset of all relations as truly found in the PW. Notice that this thus won't be a direct model of emotions. What it will be is a proof of existence of relations between emotions and the PW as well as the CO processes. It may also illustrate the change of relations. So, if we had a single receptor only in a recurrent relation with the PW what we might do is for instance look at the change of the potential as found in the potential wave for that single receptor. This is done in equation 6.6.

$$\frac{dV_w}{dt} = \frac{dV_i}{dt} \otimes V_d + V_i \otimes \frac{dV_d}{dt} \tag{6.6}$$

This gives a representation of the change of emotional state for a single receptor being. So the change of the emotional state is related to *the change* in CO desire potential, the information potential, the CO desire potential itself and *the change* of information potential. Basically all these relations are now defined. Knowing two of the three variables gives us the possibility to predict what the third one was doing. Equation 6.7 illustrates the change of CO desire. The symbol $\otimes^{-1}$ represents the inverse of the $\otimes$ function.

$$\frac{dV_d}{dt} = \left(\frac{dV_w}{dt} - V_d \otimes \frac{dV_i}{dt}\right) \otimes^{-1} (V_i) \tag{6.7}$$

Equation 6.8 illustrates the change of information potential. It isn't the same as equation 6.3 but it should be noted to be in a direct relation with that equation. What it basically does (as previously said) is partially predicting what

[3]Although connections in derivative solutions may be reversed for back propagation this has nothing to do with the generation of the current output value. Back propagation when used comes after the generation of the actual output. Back propagation cannot be used in the CRM, and thus is not suited for consciousness, because the rewards are part of the inputs. So within consciousness there has to be a learning rule which makes use of the actual inputs. Any special operations will have to be carried out within the CO process based on the given inputs.

the PW was doing. To know what the PW was doing it would also be important to know what the function of the resistance was for the receptor. What the PW was doing will then be based on a certain initial state and changes of the potential wave and CO desire.

$$\frac{dV_i}{dt} = \left(\frac{dV_w}{dt} - V_i \otimes \frac{dV_d}{dt}\right) \otimes^{-1} (V_d) \tag{6.8}$$

Something that should be stated based on the above and the fact that the PW is continually influenced by actuators which may in turn influence the being itself as well is the following.

statement 6.7 (beings and the physical world) *Beings and the physical world form a recurrent system, which means that beings are continually influencing the physical world's state and the physical world is continually influencing beings states or emotions.*

Statement 6.7 identifies the fact that beings constantly influence the PW and vice-versa. From statement 6.7 follows statement 6.8. This basically states that beings are directly or indirectly related to each other with the PW defining the relations.

statement 6.8 (associations between beings) *An associative relation exists between different beings where beings may influence each other with the physical world as a medium.*

Statement 6.8 not only defines the existence of interaction, it also clearly defines how different instantiations of the CRM may work together. By forming associations it's then possible to solve more complex problems.

You might be curious how the recurrence from statement 6.7 comes into being. To describe this we'll first have to look at what the information potential would be a function of. When it's clear and we know how the information potential behaves based on previous steps in time we can then take a look at the CO desire potential's functional behaviour. This basically means rewriting equations 6.7 and 6.8 until they contain all the variables of the previous step in time. When these two functions have been defined equation 6.6 can be used to combine the two results. Although it may just look like a whole bunch of gibberish mathematics it will serve the purpose of describing the properties of receptors and emitters well.

First let's start to derive the functional behaviour of the information potential. Transmittees come forth from activations. Activations are a function of dormants and actuators, which then may induce the transmittees. So the information potential is some kind of function of these dormants and actuators or the number of dormants and actuators as present in the PW. The PW potential V_{PW} is thus a function of the information flow exhibited by the dormants $V_{dormants}$ and actuators $V_{actuators}$. Within this function the resistance is also dependent on the dormants and the actuators and thus influences the actual flow. The dependance on the dormants and actuators is illustrated in equation 6.9.

$$\exists V_{PW}(V_{dormants}, V_{actuators}) : V \times V \rightarrow V \tag{6.9}$$

The derivative to time of equation 6.9, or the PW potential derivative, is illustrated in equation 6.10. This rule follows from the chain rule which we know from mathematics.

$$\frac{dV_{PW}}{dt} = \frac{dV_{dormants}}{dt}\frac{dV_{PW}}{dV_{dormants}} + \frac{dV_{actuators}}{dt}\frac{dV_{PW}}{dV_{actuators}} \tag{6.10}$$

The information potential or the number of generated transmittees by a receptor is a function of the PW potential. This is illustrated in equation 6.11. These may be said to be in a direct relation with each other. They cannot be identified as being exactly the same because the encoding differs. The PW potential is represented by the number of transmittees within a specific timeframe. This means that it's spread over multiple instances of transmittees. The information potential is a recording of this number in a single instance of information.[4]

$$\exists V_i(V_{PW}) : V \to V \tag{6.11}$$

The derivative to time of equation 6.11, or the information potential derivative, now becomes as illustrated in equations 6.12 and 6.13.

$$\frac{dV_i}{dt} = \frac{dV_{PW}}{dt}\frac{dV_i}{dV_{PW}} \tag{6.12}$$

$$\frac{dV_i}{dt} = \Big(\frac{dV_{dormants}}{dt}\frac{dV_{PW}}{dV_{dormants}} + \frac{dV_{actuators}}{dt}\frac{dV_{PW}}{dV_{actuators}}\Big)\frac{dV_i}{dV_{PW}} \tag{6.13}$$

The actuator potential of the given actuator emitter is a function of the CO desire. All the other actuators from other actuator emitters are now just reduced to dormants. This is illustrated in equation 6.14.

$$\exists V_{actuators}(V_d) : V \to V \tag{6.14}$$

The derivative of the equation 6.14, or the actuator potential derivative, now becomes equation 6.15. Here we should note that the CO desire potential derivative is a CO desire potential derivative from an infinitely small moment in time earlier. This has been noted by writing V_d' for the earlier step in time instead of V_d.

$$\frac{dV_{actuators}}{dt} = \frac{dV_d'}{dt}\frac{dV_{actuators}}{dV_d'} \tag{6.15}$$

Now equation 6.13 becomes equation 6.16.

$$\frac{dV_i}{dt} = \Big(\frac{dV_{dormants}}{dt}\frac{dV_{PW}}{dV_{dormants}} + \frac{dV_d'}{dt}\frac{dV_{actuators}}{dV_d'}\frac{dV_{PW}}{dV_{actuators}}\Big)\frac{dV_i}{dV_{PW}} \tag{6.16}$$

Equation 6.7 introduces the information potential as well as the potential as caught in the potential wave from an infinitely small moment in time earlier.

[4]Note that the PW potential here isn't a representative for the pressure exerted by the PW, but represents the number of generated transmittees.

This has been noted by writing V_i' and V_w' for the earlier step in time instead of V_i and V_w. Introducing equation 6.7 into equation 6.16 leads to equation 6.17.

$$\begin{aligned}\frac{dV_i}{dt} &= \frac{dV_i}{dV_{PW}}\Big(\frac{dV_{dormants}}{dt}\frac{dV_{PW}}{dV_{dormants}} + \\ &\Big(\frac{dV_w'}{dt} - V_d' \otimes \frac{dV_i'}{dt}\Big) \otimes^{-1} (V_i')\frac{dV_{actuators}}{dV_d'}\frac{dV_{PW}}{dV_{actuators}}\Big)\end{aligned} \tag{6.17}$$

Next we should describe the functional behaviour of the CO desire potential. The CO desire potential is a function of the previous CO desire as well as the information potential. This is illustrated in equation 6.18. Here the previous CO desire potential is denoted by V_d'.

$$\exists V_d(V_d', V_i) : V \times V \to V \tag{6.18}$$

The derivative to time of equation 6.18, or the CO desire potential derivative, is illustrated in equation 6.19.

$$\frac{dV_d}{dt} = \frac{dV_d'}{dt}\frac{dV_d}{dV_d'} + \frac{dV_i}{dt}\frac{dV_d}{dV_i} \tag{6.19}$$

Of course the information potential derivative, equation 6.17, can be substituted into equation 6.19. Since all the previous time-step variables are already known, I'm not going to do this. It will only provide for lengthier equations. Just as well introducing equations 6.17 and 6.19 into equation 6.6 is left up to the reader's leisure.

As you may understand modeling every single instance of a receptor and an emitter is way to complex. The ERM as proposed here, which should as stated earlier be put in a direct relation to the CRM, now is restricted to a single receptor-emitter system. I'm not going to undertake more steps towards the *complexity goal* which will eventually lead to high amounts of consciousness. I'm going to accept the model as it is.

An example of what you might predict with parts of the given model is that when you want the change of the emotion to remain constant when the information potential decreases in potential is that the CO desire will have to increase. This is nicely illustrated by equation 6.6.

What the implications are obviously depends on the particular emotion instance and the way you change the potentials. The emotion instance here would only model a very small aspect which makes up an emotion. An emotion as we know it is made up of a seemingly infinite set of these *emotion instances* or *parts of a potential wave*, depending on the contextual terminology you wish to employ. The information potential is a function of the flow and resistance. These were also influenced by the emitting of actuators which come forth from certain CO desires. As said previously this means that the relations become more and more complex and will stretch out until the beginning in time. It isn't possible to model this.

So now we basically know about the existence of emotional ties which need to be modeled. The relation to previous times will have to be implicitly caught in the model, by creating a continuous process. What the reader should also remember is that what I've said here means that for instance sensors shouldn't be simple *on* and *off* buttons, but should behave like sliders which can for

instance be *31.221* percent *on* and *68.779* percent *off*. This section mainly illustrated into what kind of behavioural model emotions should be fitted.

6.3 Emotional Expression

As we all feel, emotions aren't just experienced in our heads, right? We feel the beating of our hearts and have over a long period of time even attributed this beating the honour of being the centre of our souls. How wrong we were. The reason why we bestowed upon our heart the honour of being the centre of emotions is because our emotions are directly expressed in certain sensations interior to our body as well as exterior representations. Or basically the emotions are expressed in the PW.

This section will describe briefly how this happens. It follows directly from the CRM, so it won't be a long section.

What actually happens is that the emotion triggers certain physical reactions by sending signals to the PW. Seen from a relational point of view this could be described as in statement 6.9.

statement 6.9 (emotional expression) *The emotional expression builds a physical representation of the emotion in the relations a being is directly associated with.*

Seen from the CRM an emotion which is represented by the potential wave is processed by the CO process. This expresses a certain CO desire, which not only partly induces the next emotion in a direct manner, but also induces the release of actuators. These actuators cause the instantiation or initiation of a process which will physically represent the emotion. The physical representation may then be perceived by us or other beings.

statement 6.10 (expressing emotions) *The process of expressing emotions is initiated by the emitting of actuators.*

What needs to be perceived here is the difference between expressing emotion and the language of expressing emotions. Why is it that we associate a mouth that turns the corners up with a happy emotion and a mouth that turns the corners down with a grumpy emotion? Evolutionary pre-programming and culture are the most likely causes, because there's no really apparent reason. The amount of joy within a robot might just as well be indicated by a metre to the outside world, as well as by not representing it at all.[5] A seemingly empty shell need not be empty.

The expression of an emotion is greatly determined by the possibility of expressing emotions, just as well as by a being's cultural heritage. For instance, our representation of laughter may be completely misunderstood by beings that do not share our language of emotions. It isn't wise to grin at a gorilla for as far as I know.

definition 6.11 (emotional language) *The emotional language is the specific instantiation of physically induced actions which represent a certain internal state of a being.*

[5]Of course there's no such thing as an emotion representable by a joy-metre, because joy is a pre-classified emotion and so are other emotions. A couple of metres will therefore never be able to capture the whole of the emotions.

statement 6.12 (emotional language) *Interpretation of emotions depends on the preconceived emotional language.*

What does make the distinction between the expressions then? Funnily, when you're not happy, your mother will probably tell you to put on a smile. This smile will then provide you with a positive stimulus which may actually initiate a happy mood. People with attitude come a long way. Attitude comes forth from boosting your own ego.[6]

The emotions are basically represented by equal reactions that they induce in our internal physique. This would to my opinion be the best way to model them. Don't tell a robot to smile, let it discover its smile. This is already an instantiation of a very specific system, so I'm not going to explore this further here.

statement 6.13 (representation of emotional expression) *An emotional expression is meant to induce the perception of that what it represents.*

6.4 Existing Models of Emotion

What is now the main question is whether current models implement these relations that I have defined. Most of these models implement certain BES and CES. The main problem is that emotions are experienced to certain degrees. I can be happy to a certain degree. I can be a little bit sad. Because of this, systems that only implement the representation or utterance of extremes in emotions don't appear to be natural. Just as well, the changing of emotional states happens gradually and not at once.[7] These emotions also change in accordance with experiences of the PW. When these emotions don't change naturally, people will not accept it. If you're just going to model emotions and not a conscious system, this will lead to the need to define emotional states. Without consciousness no proper interaction can take place between beings which leads to the expression of emotions. Even the most simple system which has emotions, like for instance a tutoring system, should be aware of the reactions evoked by its emotions. True emotions are dynamically reactive systems. In order to do this learning should take place constantly. In order to properly deal with the environment the CRM should explicitly or implicitly be present. Although I am against it seen from the point of view of a conscious representation, as a separate model ignoring the CRM I cannot but accept classifying emotional models for what they are.

Examples of models to describe emotions are for instance fuzzy systems. One such system is called FLAME.[8] The FLAME model in the paper used to present it was compared to a strict rule-based system and a random emotion system. The FLAME model gave people the idea that the program responded life-like. It implemented a learning mechanism which adapted the being's emotional response based on past experience. It learned based on current emotional response and the PW influence. Unfortunately these kinds of systems have certain limitations.

[6](Kall, 1989)

[7]This may according to our perception still happen instantly where *instantly* means that it happens thus fast that we can't perceive it happening.

[8](El-Nasr and Yen)

The limitations that these systems have are imposed on them by their designers. The designers still have to define certain action-response mechanisms. The actions and responses may be divided into more fuzzy groupings and expressions of these groupings, but they are still predefined. A problem that's associated with this is the number of rules to create a believable actor.[9] Here the actor's facial expressions are controlled by fuzzy mechanisms. The number of rules in the implementation discussed in the paper approximates *10,000* (which only partly leads to the wanted natural result). As you may understand, there must be a better way to do this.

Although there is a better way to do this[10] this will still be a fuzzy system. Our emotions change gradually. Since it should be a fuzzy system, apparently our conscious process implements all of the fuzzy system's properties. How does it do this? Or to generalise this: since the GToC should lead to consciousness, how is the fuzzy architecture represented by the GToC?

What does a fuzzy system contain? A fuzzy controller contains four components.[11] They are:

- **Fuzzy rule base:** The rule base, or knowledge base, contains the fuzzy rules that represent the knowledge and experience of a human expert of the system. These rules express a nonlinear control strategy for the system.

  While rules are usually obtained from human experts, and are static, strategies have been developed that adapt, or refine rules through learning using neural networks or evolutionary computing.[12]

- **Condition interface (fuzzifier):** The fuzzifier receives the actual outputs of the system, and transforms these non-fuzzy values into membership degrees to the corresponding fuzzy sets. In addition to the system outputs, the fuzzification of input values to the system also occurs via the condition interface.

- **Action interface (defuzzifier):** The action interface defuzzifies the outcome of the inference engine to produce a non-fuzzy value to represent the actual control function to be applied to the system.

- **Inference engine:** The inference engine performs inferencing upon fuzzified inputs to produce a fuzzy output.[13]

One thing that I cannot agree with is that a fuzzy rule base has to necessarily represent a human's knowledge base. We humans are fuzzy systems in the way we work, but we were programmed by evolution. This is fortunately already addressed by referring to evolutionary programming. The outcome of evolution can be modeled by a fuzzy system, but this outcome will need to take its place in evolution as well. Basically us using evolutionary programming is already an outcome of evolution which leads to the implementation of a fuzzy system. It's just that the process of creating such systems and the further life of these systems should remain dynamical regarding procreation and evolution. This

[9] (Karunaratne and Yan, 2000)
[10] This we can be sure of, because we ourselves exist.
[11] (Engelbrecht, 2002)
[12] (Favilla et al., 1993; Wang and Mendel, 1992)
[13] (Engelbrecht, 2002, Section 19.2)

can only happen when the thing created is properly evolved within our own world.

The *fuzzy rule base* is the ever changing collection of rules that makes sure beings survive and evolve as a species. The rules as imposed on a being by its social position determine the way it lives and survives. These rules are basically derived sub-goals of the goal of survival: the continuation of the process of life. Are there any pre-programmed rules in evolution? No. Although we call it *survival of the fittest*, this is just a particular instance. What evolution really is is captured in definition 6.14. There are no true explicit rules. There are only emergent implicit rules.

definition 6.14 (evolution) *Evolution is the collection of continuations and cessations of processes.*[14]

What determines what process is better is actually the fact that one process continues and the other doesn't. Although usually estimates will turn out to be true regarding which process is better, this may not always be so. Sometimes a process with a rather high likelihood to fail succeeds due to circumstances. So there is no other function other than the continuation and cessation of processes determining which process at a certain point in time is better.

The *condition interface* is provided by the signals that are received from the PW and the actual CO desires that the CO process produces. The membership degrees are represented by the potentials caught in the potential wave. This means that the output of the condition interface is always directly linked to the input of the condition interface, capturing its last output state. The thing that changes the CO state is the CO state itself as well as the PW.

The *action interface* is represented by the process of emitting actuators into the PW. These actuators will induce certain processes which may lead to actions. This may mean the stimulus of certain receptors, where the actuators act as dormants becoming activators. Just as well the actuators may induce a certain potential over muscle areas, stimulating the muscle to contract. An actuator may have many functions.

The *inference engine* is basically the CO process which produces certain CO desires, where the CO desires receive certain potentials representing their fuzzy values. The fuzzy outputs are on the one hand made known in the form of caught potentials in the potential wave. On the other hand the fuzzy values are propagated to the outside world. By emitting a certain number of actuators during a certain time-frame these fuzzy values are transformed into non-fuzzy values.

So what does the model I propose have what the models that already have been put in use haven't? Basically the model I propose is fully adaptive to its position in the world. It adjusts to the relations around it as a means to continue its process. It tries to fit within the relations that make up the world without speeding up the cessation of its own process or that of its species.

[14]This isn't an empty definition. It's actually the most general definition of evolution taking into account all relations. From this may be derived fitness functions which go way beyond who's the strongest. Contrary to former definitions of evolution it leaves no room for discussion. It's a closing definition. Because of the high abstraction level it may seem empty, but when removing everything that's largely situation dependent, this is all that's left. So this is the only general definition of what evolution is which applies to all applications and implementations.

From a relational point of view a conscious process is made up of smaller conscious processes. This accounts for individualism as well as conformism to be part of a single chaotically coherent structure called society. Just as well it accounts for the smaller processes running in a being working together to form that one being. This way both a single instance of a being as well as a group of these instances may be seen as a single being.

The models that were proposed in the past had a predefined world with predefined rules. When the model was put outside of its predefined world it wouldn't be able to survive. Of course when a natural being is removed from its habitat without providing artificial support it also fails to live. This is inherent to the nature of being. But what I would like to propose is a more adaptive model that can be fitted into all different sorts of societies, where it isn't dependent on the CO structure, but on its representational architecture.

The CRM as I propose it deals to a large extent with this problem, because it lacks a predefined architecture. The processes of any being can be represented by the CRM, where the expressions are defined by the architecture. This can be done by letting it discover its expressions as I stated previously. By letting it discover its expressions the act of expressing will become more natural as well. It may not only start using its expressions to express its moods, but also to change its own mood. To do this a mechanism should be proposed to let it learn its emotional mechanisms.

To a certain degree there may also be reflexes directly coupled to the expressions. These will run via different paths as they do in beings as we know them. But these shouldn't be the only mechanisms.

An example of such reflexes is the laughing reflex with some patients whose faces have been partially paralysed. When you ask such a patient to laugh, only half of the face will. When you make such a patient laugh, the whole face will laugh. This is due to the fact that the motor cortex has been disconnected, which can normally be controlled by us to induce a smile. This is also called the upper nerve pathway. But when the lower nerve pathway which is directly linked to the part of our brain where our emotions are mediated is still intact, we may still reflexively smile.[15]

The whole of the emotion as we experience it doesn't only come forth from the part where the emotion is mediated. It usually also contains a certain associative content which the rest of our cognitive process can relate to. And if not, an emotion that hasn't been experienced before may for instance frighten us, even though it may actually be a normal good feeling. It may also let us react in some other strange way, because of the lack of definition of the association. The main problem with emotions as they have been modeled is that they behave according to rules and thus will never rise above being a reflexive system.

6.5 Mechanisms for Emotions

Emotions are very complex. This means that some basic mechanisms should be identified which need to be implemented in a being in order to properly deal with emotions. In the previous sections numerous issues were identified, but there are still some things that need to be dealt with. In order to so so this

[15](Kall, 1989)

section describes the basic instantiations of the CRM and their functionality regarding emotions.

In order to deal with emotions a being needs to be able to identify its emotions. Before a being can do anything with its emotions it must first have a clear view of what emotion it's actually having. The emotion a being is having may be directly derived from its complete physical state, using mostly its senses. This comprises tension, movement, heat, cold and all the other aspects that make up emotions.

A being needs to be able to express its emotions. The emotions a being has must be directly representable by the physical properties of a being. So a being doesn't only sense its physical properties, giving the being a clear picture of its emotions, but it also induces them.

A being must be able to perceive the rewards associated with the emotions. When emotions are expressed these expressions will feel good or bad. The rewards associated with the emotions are basically the same instances as everything that's used to identify the emotions. The difference lies in what is done with the emotions. Based on its perception of emotions it not only takes certain actions, but it also learns. Of course perception and learning are both caught in the CRM, so in case of instantiations of the CRM it isn't necessary to implement anything extra for the reward mechanisms.

A being needs to be able to control its emotions. The control of the emotions means that the being will try to change the physical make-up of its emotional state. So when it feels its heart is beating too fast it may decide to calm down. This means relaxing its body entirely by for instance breathing. This will then change its physical state.

A being needs to be able to influence the physical expression of emotions. In some cases the being may not wish to express its felt emotions. This means that it will basically try to influence its exterior physical make-up visible to others. This is basically the same mechanism as the one influencing the emotion itself, it will just work on a smaller scale.

When all these mechanisms have been implemented which will need a suitable morphology (embodiment) for a being, it's then possible to actually implement or discuss such a being. Without a proper morphology emotions are to complex and too subtle to model.

Chapter 7

Foundation

What's a theory without proper foundation? Based on what may we decide that this theory is correct? *Theories are not verifiable, but they can be 'corroborated'.*[1] So, what can be *corroborated*? The theory states that if one of the basic processes isn't implicitly or explicitly there, then there's no way that any kind of consciousness may emerge. Furthermore these processes have already been put in a specific order. So, to start with, the presence and order of these processes should be tested based on what we perceive to be our knowledge. A major question might then still be whether the theory is complete. This would mean checking whether there shouldn't be another process which is necessary for consciousness to emerge. At first the foundation laid in this chapter was supposed to be based on psychological evidence. This turned out not to be necessary.

Section 7.1 discusses the necessity of each of the layers. Section 7.2 discusses what would happen if the arrangement of the layers was changed. Section 7.3 discusses the completeness of the theory.

7.1 Presence of the Layers

In order to discuss the necessity of the layers it's easiest to look at what would happen if each layer wasn't there. So basically for each of the seven layers the effects are here described regarding what would happen if the layer wasn't there. The process propagates up and down. This will also be the order in which it is discussed.

The PW layer needs to be there in order to have something to perceive[2] and in order to have a structure (a medium) through which dormants may be propagated.[3] Suppose the PW layer wasn't there. Now there would be nothing to perceive. Just as well there would be no medium to send the dormants and actuators over or through. For instance light propagates through space. Another example might be that pulses of current are propagated through the spaces between axons and dendrites of neurons.

[1] (Popper, 2004)

[2] What is perceived are then dormants, including its derived subclass actuators.

[3] This may also be emptiness as a medium.

The TR layer needs to be there in order to receive dormants. Suppose the TR layer wasn't there. Without the TR layer there's no way to determine what the current PW state is. It's not possible to receive any dormants to process. For instance a blind man can't see. Just as well a neuron without dendrites has no inputs. In case of social consciousness for example someone without a computer may not receive its E-mail.

The NE layer needs to be there to connect different processing units as well as to define the topography of the NE. Suppose the NE layer weren't there. On the one hand it wouldn't be possible to send transmittees from a receptor to the CO process in the CO layer. Just as well it isn't possible to define the LO of the emitting of transmittees. So it isn't possible to determine where the original signal captured in the transmittee came from. For instance in case of social consciousness people without anything like a phone-line, E-mail or any other kind of (maybe indirect) communication means cannot communicate over long distances. So it isn't possible to socially connect in some way. In case of individual consciousness neurons which haven't been connected with each other in the least amount will not work together.

The LO layer needs to be there to attribute a LO to a transmittee. Suppose the LO layer wasn't there. Now it isn't possible to attribute a LO to the transmittee. So it isn't possible to define whether a stimulus was a sound or a smell or maybe light or touch. Everything is just one big stimulus. Just as well it isn't even possible to determine at what time the original stimulus arrived. So quite possibly a stimulus may actually have been perceived two hours late.

The DS layer needs to be there to collect the transmittees into information instances. Suppose the DS layer wasn't there. Transmittees are then sent up, meant for a CO process, but they aren't collected. So the CO process doesn't do anything with the transmittees and thus cannot do anything at all.

The RE layer needs to be there to build the input for the CO process in the form of the potential wave. Suppose the RE layer wasn't there. Now the information instances are actually collected, but the CO process' requirements aren't met. So the CO process cannot do anything with the information instances because they aren't the type of inputs that it wishes to receive. So the CO process isn't able to do anything.

The potential wave that is built also takes the desire uttered by the CO process into account. Suppose this wasn't done. A single processing unit would then not be able to remember its own state. This is however needed to iteratively walk through a few steps of reasoning. On the one hand the GToC then wouldn't model social consciousness, because for instance the humans making up a social NE wouldn't be able to reason about anything. In case of individual consciousness, modeling single neurons, this would mean that the neuron wouldn't try to maintain a certain average state it was in. So in case of a neuron the desire which is uttered is partially implicitly modeled. On the one hand there's the output that's generated by the axon. On the other the neuron has a certain physical state.

The CO layer contains the CO process which forms the knowledge base. Suppose the CO layer isn't there. Then there's no process generating the desire as output. There's no knowledge base. There's no way to do anything with the incoming signals. So there's no way to be conscious. In case of social consciousness this would for instance mean reducing some being's brain activity to nothing. In case of individual consciousness it would for instance mean

removing everything from the neuron that's between the dendrites and the axon. Now no intermediate process could take place.

The RE layer also needs to be there to handle the output of the CO process in terms of the uttered desire. Suppose the RE layer wasn't there. Then the desire uttered by the CO process would not be expressed, nor would it be part of the input for the CO process (the potential wave). So the CO process would have no way to for instance remember its own current state in reasoning. This may then of course again be a physical state in case of a neuron or it may be non-physical if we see ourselves as processing units in a social NE.

The DS layer needs to parse the output of the CO process from the desires into separate signals. Suppose the DS layer wasn't there. The output of the CO process, i.e. the desire, would now not be understandable by the processes which lead to the emitting of actuators. So the output needs to be parsed *from*, just as well as the input data needed to be parsed *to*, the required form for the CO process. The type of data which is generated from the desires is a set of information instances.

The LO layer sends the signals contained in the information instances to the right LOs as transmittees. Suppose the LO layer wasn't there. This way the transmittees would be sent without knowing where they would arrive. So it's never possible to deliberately influence anything, because you don't know what influence will be expressed where. You would just be sending transmittees to random LO hoping that they might also arrive for some part at the desired LO. This would in case of individual consciousness for instance result in spasms. It may now also happen that a stimulus is uttered five seconds later than should have been the case. It's a bit like sending an E-mail which says "The bomb is exploding in three hours and fifteen seconds!" without a time-stamp. You can never be sure how much time has already passed.

The NE layer needs to be there to connect to the different emitters and thus also still defines the topography of the NE. Suppose the NE layer wasn't there. Then it wouldn't be possible to emit anything, because there was no connection to the emitter.

The TR layer needs to be there to emit actuators into the PW layer. Suppose the TR layer wasn't there. Now there wouldn't be any emitters and no possible influence can be expressed into the PW layer. Regarding individual consciousness this would mean trying to move a limb and not being able to even though you've got all the room you need. Regarding social consciousness this would mean not being able to express you opinion because your mouth is being held shut.

The PW layer receives these actuators and processes them. Suppose the PW layer wasn't there. In order to have an influence on anything, this anything needs to be there. If it isn't there there's for instance nothing possible to explore. Just as well it isn't possible to send any findings to another processing unit.

7.2 Arrangement of the Layers

The hard way to do this is discussing any arrangement of the seven layers. This would mean describing 5040 (= 7!) different arrangements. In order to take a shortcut this section will only look at the adjoining layers in the CRM and discusses why each of the layers should be above the other. Since this holds for

all layers, the present arrangement of the layers is then preserved.

The TR layer is on top of the PW layer. Suppose the TR layer was lower than the PW layer. The TR layer now has no physical medium to receive dormants from. Nor may it emit actuators onto any such medium, because it isn't there. So the TR layer needs a medium which it finds in the PW layer. The TR layer should thus be above the PW layer.

The NE layer is on top of the TR layer. Suppose the TR layer was above the NE layer. The NE layer defines the topography of the connections of the receptors and emitters to the CO process. When the TR layer is now said to be above the NE layer the receptors and emitters have no connection with the CO process. So the NE layer needs to be in between the TR layer and the CO layer. This means that it has to be above the TR layer.

The LO layer is on top of the NE layer. In order to define the LO of a signal you need a certain topography. When the NE layer and the LO layer are switched the LO cannot be determined because the reference to the LO in space isn't fixed. Let's illustrate this. You can say your from Earth or Mars, but these names mean nothing without a clear perception of what Earth and Mars are. Just as well it's possible for two entities to claim that they are from Mars even though they originate from totally different places. Since there's no topography, the two can't be proven to be contradictory regarding their statements. Without a topography there's no location. So the LO layer should be above the NE layer.

The DS layer is on top of the LO layer. If the LO layer would be above the DS layer, this would mean that the transmittees wouldn't have a LO associated with them. This would also mean that there's no possibility to gather the transmittees as information instances per each receptor. It will not be possible to attribute to every type of information a certain heaviness of the incoming signal. This means that the transmittees will need a certain LO before they may be gathered. So the DS layer needs to be above the LO layer.

The RE layer is on top of the DS layer. Suppose the DS layer were above the RE layer. First of all the RE layer wouldn't have any collected data to represent. Second, if it would represent something the DS layer would be able to alter this RE. This would mean that the data would be distorted and the possibility exists that the CO layer doesn't understand it anymore.

The CO layer is on top of the RE layer. Without a proper RE of the incoming data which the CO layer can handle, the CO layer can't do anything. So the CO layer needs to be above the RE layer.

7.3 Completeness

What has been proven is that every layer which is present now should be present. This basically fulfills the statement that this is the abstract definition of the process required for the emergence of consciousness in so far that these layers need to be present.

Is it possible to prove that there should be more? This can only be proven by coming up with something that's still missing and isn't modeled in the theory. Based on what we know the likelihood for this to happen is ultimately low, especially considering the fact that even qualia[4] fit into the picture of the GToC. It isn't possible to prove the completeness of the model. It is however by far

[4](Hobo, 2004b, Section 2.3)

the best model which has been provided, especially since it's a closing model regarding its definitions.

Chapter 8

Conclusions and Recommendations

This thesis identified several goals which had to be met in order to gain a better understanding of consciousness. Understanding consciousness better doesn't mean that we can now stop thinking. What should be done with the knowledge gained is also an important research question. First a discussion is held for each of the goals to determine whether they have been met. Then recommendations are made regarding the work which should be done in the future.

First of all this thesis proposes an abstract architecture to describe the basic processes which need to be present for consciousness to emerge. The abstract architecture is clearly defined and illustrated.

The model applies to any type of being. If there is a being which the model does not describe, this being will still have to be found. This being will then not *not contain* these layers, but it will contain more. It is of course not proven that we don't contain more, but it is highly unlikely to be so.

For different research areas this model also has to form a proper foundation to speak about consciousness. The definitions are formed in such a way that it's easily comprehensible in which stage of consciousness these processes reside. There's no possibility other than not knowing the model or not referencing the model properly in order to mix up the definitions.

The functionality is clearly defined. Every function has a clear input and output specification defining its behaviour.

The model is clear enough to serve as proof for the succeeding or failing of theories. When a theory doesn't stand up, cannot be generalised in order to adhere to the model, it lacks functionality. The theory is incomplete and thus will not work. Of course when only speaking about parts of consciousness it may only need parts of the functionality.

The model is a general model which forms a clearer basis for interaction. The way relations come into being between different processing units in case of social consciousness is clearly defined. For instance agents designed for interaction may now be designed thus that they behave more naturally. The proper forming of the relations makes sure that the agents behave in a natural manner.

The possibilities we will have in the future will probably exceed those we have now. There's thus one more thing that should be considered. How far

should we go? To start this discussion I would also like to propose a discussion about ethics. These ethical rules should be concerned with what is acceptable in dealing with organisms (artificial or non-artificial). A first start has been made in my Master's thesis appendix I.[1] Hopefully the work done there will also grow a better understanding of what is actually happening in doing some of the experiments that are actually already done today.

The goals have been properly met in the thesis. This doesn't change that a lot still has to be done. To my opinion not only ethics should be derived, but also from these ethics we should conclude what is desirable to implement. It may for instance in many cases be possible to implement simplified models which predict behaviour properly and may be used to give us the idea that proper interaction is taking place. These simplifications then would not be conscious. Given the GToC these simplifications should be easily derivable.

[1] (Hobo, 2004a)

Bibliography

D.J. Chalmers. *The Conscious Mind: In Search of a Fundamental Theory*. PHILOSOPHY OF MIND SERIES. OXFORD UNIVERSITY PRESS, New York, Oxford, 1996. ISBN 0-19-511789-1.

P.S. Churchland. Can neurobiology teach us anything about consciousness? In N. Block, O.J. Flanagan, and G. Gülzedere, editors, *The Nature of Consciousness: Philosophical Debates*. The MIT Press, Cambridge, Massachusetts, 1997. ISBN 0-262-52210-1.

A. Einstein. The foundation of the general theory of relativity. In A. Sommerfeld, editor, *THE PRINCIPLE OF RELATIVITY a collection of original papers on the special and general theory of relativity*. Dover Publications, Inc., 1952.

M.S. El-Nasr and J. Yen. Agents, emotional intelligence and fuzzy logic. Computer Science Department, Texas A&M University.

A.P. Engelbrecht. *Computational Intelligence: An Introduction*. John Wiley & Sons, Ltd., 2002. ISBN 0-470-84870-7.

J. Favilla, A. Machion, and F. Gomide. Fuzzy traffic control: Adaptive strategies. In *IEEE Symposium on Fuzzy Systems*. San Francisco, 1993.

O. Flanagan. Prospects for a unified theory of consciousness or, what dreams are made of. In *The Nature of Consciousness*. The MIT Press, Cambridge, Massachusetts, 1997. ISBN 0-262-52210-1.

S. Handel. *Listening: An Introduction to the Perception of Auditory Events*. MIT Press, 1989.

Emile Michel Hobo. (appendix i) a general agent design specification. Master's thesis, University of Twente, 2004a.

Emile Michel Hobo. (appendix ii) derivative ideas and considerations based on the general theory of consciousness. Master's thesis, University of Twente, 2004b.

J. Jacky. *The way of Z, practical programming with formal methods*. Cambridge University Press, 1997. ISBN 0-521-55976-6.

R. Kall. Emotional self regulation and facial expression muscle measurement and training. In J. Cram, editor, *Clinical Surface EMG*, volume 2. 1989. downloaded on March 8^{th}, 2004, http://www.futurehealth.org/SmileAnatomy.htm.

S. Karunaratne and H. Yan. A fuzzy rule-based interactive methodology for training multimedia actors. In *Selected papers from Pan-Sydney Workshop on Visualisation*, volume 2 of *ACM International Conference Proceeding Series*, pages 3–9. Australian Computer Society, Inc., Darlinghurst, Australia, 2000. ISBN/ISSN: 1445-1336 , 0-909-92580-1.

R. Manzotti. *Intentional robots - The design of a goal-seeking, environment driven, agent*. PhD thesis, LIRA Lab, DIST, University of Genoa, 2003a.

R. Manzotti. A process based architecture for an artificial conscious being. *Axiomathes*, September 9^{th} 2003b. Riccardo Manzotti is associated with: LIRA Lab, DIST, University of Genoa.

A. Ortony, G.L. Clore, and A. Collins. *THE COGNITIVE STRUCTURE OF EMOTIONS*. Cambridge University Press, 1999. ISBN 0-521-38664-0.

R. Penrose. *Shadows of the Mind - A Search for the Missing Science of Consciousness*. Oxford University Press Inc., New York, 1994. ISBN 0198539789.

R. Pfeifer and C. Scheier. *UNDERSTANDING INTELLIGENCE*. Massachusetts Institute of Technology, 1999. ISBN 0-262-16181-8.

K. Popper. *The Logic of Scientific Discovery*. Routledge Classics. Routledge, London and New York, 2004. ISBN 0-415-27844-9.

A. Sloman. *THE COMPUTER REVOLUTION IN PHILOSOPHY: Philosophy Science and Models of Mind*. THE HARVESTER PRESS LIMITED, Sussex, 1978. ISBN 0-85527-542-1.

A. Sloman. A systems approach to consciousness (how to avoid talking nonsense?). Aaron Sloman is associated with the School of Computer Science, The University of Birmingham, 1996.

A. Sloman. Architectural requirements for human-like agents both natural and artificial: What sort of machines can love? In K. Dautenhahn, editor, *Human Cognition and Social Agent Technology*, chapter 7. John Benjamins Publishing Company, Amsterdam/Philadelphia, 2000. ISBN 90 272 5139 8 (Eur.) / 1 55619 435 8 (US).

A.S. Tanenbaum. *Computernetwerken*. Academic Service, Schoonhoven, 2^{nd} edition, March 2000. ISBN 90 395 0557 8. Original title: Computer Networks Third Edition.

The Alzheimer Association. The alzheimer's association. http://www.alz.org/, February 2004.

C.A. Vissers, L. Ferreira Pires, D.A.C. Quartel, and M.J. van Sinderen. *Ontwerpen van telematicasystemen*. University of Twente, Department of Computer Science, Enschede, the Netherlands, March 2002. Vakcode 214005.

L.X. Wang and J.M. Mendel. Generating fuzzy rules by learning from examples. In *IEEE Transactions on Systems, Man and Cybernetics*, volume 22, pages 1413–1426. 1992.

Samenvatting

Het bewustzijn is één van de weinige direct beschikbare maar moeilijk te doorgronden principes waarmee wij bekend zijn. Dit verslag introduceert een architectuur welke het bewustzijn probeert te onderbouwen. Het verschil met voorgaand onderzoek naar het bewustzijn is dat er ten eerste een (sluitend) abstract model aangeleverd wordt en ten tweede ambiëert dit model ook in hogere mate een volledig model te zijn. Het model komt voort vanuit de ingeving dat elk type bewustzijn gemodelleerd kan worden aan de hand van netwerken. De architectuur is dan ook direct ontleend aan een algemene netwerkarchitectuur. Op basis van dit model zijn er vervolgens verdergaande onderwerpen te bespreken.

Hoofdstuk 1 geeft weer wat de doelen en de aanpak van dit afstudeerverslag zijn.

Er wordt een abstracte architectuur geleverd. Deze beschrijft elk mogelijke type wezen. De architectuur dient als basis voor verschillende vakgebieden om samen te komen tot oplossingen van problemen. Door de functionaliteit beter te begrijpen is het ook mogelijk de benodigdheden van de onderliggende invulling van het model beter te begrijpen. Het model dient ook als een beter interactiemodel voor algemene applicaties.

Dit afstudeerverslag is iteratief geschreven door steeds weer te kijken naar wat er al gedaan is en dit in het afstudeerverslag te verwerken. Het afstudeerverslag is in de volgorde van de hoofdstukken gegroeid.

Hoofdstuk 2 introduceert de lagen van de architectuur. Ook laat dit hoofdstuk zien waar de lagen vandaan komen. De architectuur heet het *Consciousness Reference Model (CRM)* en is afgeleid van het *Open System Interconnection Reference Model (OSI RM) (Tanenbaum, 2000)*. Het CRM bestaat uit zeven lagen:

1. De *Physical World (PW) layer* bevat de signalen en relaties zoals die binnen de fysieke wereld worden doorgegeven.

2. De *Transmission (TR) layer* vormt het interface tussen de fysieke wereld en de verwerkingsarchitectuur van een wezen.

3. De *Network (NE) layer* vormt de transport architectuur tussen verschillende verwerkingseenheden van de gehele architectuur alsmede tussen de fysieke wereld en de architectuur.

4. De *Locality (LO) layer* geeft weer vanaf of naar welk interface de signalen worden gepropageerd.

5. De *Data-flow Selection (DS) layer* verzamelt de relevante signalen voor een gegeven verwerkingseenheid.

6. De *Representation (RE) layer* construeert de invoer voor de *CO layer*.

7. De *Cognitive (CO) layer* bevat het feitelijke verwerkingsproces.

Hoofdstuk 3 bespreekt hoe de processen precies gemodelleerd gaan worden. Dit gaat gebeuren aan de hand van netwerk modelleringstechnieken.

Hoofdstuk 4 geeft nog kort even aan wat er qua literatuur onderzoek nog is gedaan wat niet al (voldoende) toegelicht werd in de rest van het verslag. Dit ondersteunt niet alleen mijn doel, maar ook mijn aanpak. In dit hoofdstuk wordt ook een nieuwe wetenschappelijke procesgang voorgesteld. Volgens deze kan na het afleggen van die procesgang het wetenschappelijk proces geanalyseerd worden. Er is echter geen vaste wetenschappelijke methode die van tevoren aangenomen en gehanteerd kan worden.

Hoofdstuk 5 bespreekt de invulling van de processen van elk van de lagen. De signalen zoals doorgegeven van de fysieke wereld aan een wezen heten *dormants*. Deze worden doorgeven via *receptors*. De signalen gegenereerd door een wezen heten *actuators*. Deze worden doorgegeven via *emitters* aan de fysieke wereld. De *receptors* en *emitters* vormen de interfaces met de fysieke wereld. De interne communicatie wordt geregeld door het doorgeven van *transmittees* over het netwerk. Voor elk van de verwerkingseenheden worden de *transmittees* van een zeker tijdsframe verzameld in *information* instanties voor elk van de afzonderlijke *receptors*. Samen met de voorgaande uitvoer van het cognitieve proces, genaamd *desires*, wordt een nieuwe invoer geconstrueerd voor het cognitieve proces in de vorm van de *potential wave*. Dit proces genereert dan weer een nieuwe uitvoer en past zijn eigen proces aan aan de hand van gegeven invoer. De uitvoer wordt niet alleen als nieuwe invoer gebruikt, maar induceert ook weer een proces naar de fysieke wereld. Dit proces dient er dan toe om *actuators* te genereren. Deze kunnen dan de fysieke wereld alsmede de waarneming beïnvloeden. Ze kunnen eventueel ook zelf waargenomen worden als er bijbehorende *receptors* bestaan.

Hoofdstuk 6 beschrijft de eigenschappen van de basis functionaliteit welke tot emoties kan leiden in de vorm van het *Emotional Reference Model (ERM)*. Dit beschrijft aan de ene kant de emotionele inhoud zoals die binnen het *CRM* gerepresenteerd wordt. De emotionele inhoud zit bevangen in de fysieke eigenschappen van het wezen zelf in relatie met zijn cognitieve verwerkingsproces en wordt dus gerepresenteerd in de *potential wave*. Na dit vastgesteld te hebben beschrijft dit hoofdstuk ook het basisgedrag van het *CRM* aan de hand van differentiaal vergelijking. Dit definiëert ook nog eens expliciet de aanwezigheid van een relatie tussen wezens en de fysieke wereld. Tevens geeft dit ook duidelijk aan dat we in het geval van wezens niet te maken hebben met simpele *aan* en *uit* knoppen, maar met variabele *reële* invoerswaarden. Vanuit deze waarden ontstaan dan zekere expressies op basis van hoog complexe samenstellingen van processen.

De geuite emotie kan direct in relatie gesteld worden tot de emotie zelf. Deze wordt niet alleen geïnduceerd door de emotie, maar de emotionele uitdrukking induceert zelf ook in zekere mate de emotie. Dit kan natuurlijk allemaal in relatie gesteld worden tot voorgaande emotie-modellen die over het algemeen falen met het classificeren van *Basic Emotional States (BES)* en de daaruit samengestelde

Composite Emotional States (CES). Het CRM classifiëert niet. De emoties komen tot uiting door een *fuzzy model*. Het *CRM* bevat de functionaliteit van zo'n *fuzzy model*. De reflexen die tot uiting komen voor wat betreft emoties kunnen ook gemodelleerd worden aan de hand van het *CRM*.

Tot slot van het hoofdstuk worden er ook nog mechanismen voorgesteld die benodigd zijn om tot emoties te komen. Een wezen moet emoties kunnen (leren) identificeren. Een wezen moet emoties kunnen (leren) uiten.

Hoofdstuk 7 beschouwt de onderbouwing van de theorie. Hieruit blijkt dat het niet mogelijk is een van de lagen weg te laten, noch de volgorde van de lagen te veranderen. Voor wat betreft de compleetheid van het model kan geen garantie gegeven worden dat er niet nog een laag dient te zijn, maar tot deze laag gevonden wordt kan dit model als compleet beschouwd worden. De waarschijnlijkheid dat er nog een laag gevonden wordt, gebaseerd op onze inschatting van hoe alles werkt (onze "kennis"), is zeer gering.

Conclusies en Aanbevelingen sluiten de theorie af. In conclusie kan gesteld worden dat de doelen behaald zijn. Er dient nog wel uitgebreid nagedacht te worden over ethische consequenties. Vanuit deze ethische consequenties kan dan ook besproken worden in hoeverre architecturen een simplificatie dienen te zijn van het bewustzijn. Een dergelijke simplificatie houdt dan in dat de instantiatie van zo'n architectuur geen grenzen overschreidt voor wat betreft ethiek gerelateerd aan het bewustzijn.

Summary

Consciousness is one of the few directly accessible but hard to understand principles we are familiar with. This thesis introduces an architecture which tries to substantiate consciousness. The difference with previous consciousness research is that firstly a (closing) abstract model is provided and secondly this model aspires to be to a high extent a complete model. The model comes forth from the hunch that every type of consciousness can be modeled using networks. The architecture has in principle been derived from a general network architecture. Based on this model different subjects may be discussed.

Chapter 1 represents what the goals of and approach to writing this thesis are.

The thesis proposes an abstract architecture. This describes any type of being. De architecture serves as a basis for different research areas to solve problems more properly together. By understanding the functionality better, it's also possible to understand the necessities of the underlying implementation of the model better. The model also serves as a better interaction model to general applications.

This thesis has been iteratively written by repeatedly looking at what has already been done and processing this in the thesis. The thesis has grown in the order of the succeeding chapters.

Chapter 2 introduces the layers of the architecture. This chapter also shows where each of the layers comes from. The architecture is named the *Consciousness Reference Model (CRM)* and has been derived from the *Open System Interconnection Reference Model (OSI RM) (Tanenbaum, 2000)*. The CRM consists of seven layers:

1. The *Physical World (PW) layer* contains the signals and relations as they are passed within the PW.

2. The *Transmission (TR) layer* forms the interface between the PW and the processing architecture of a being.

3. The *Network (NE) layer* forms the transport architecture between the different processing units of the entire architecture as well as the PW and the architecture.

4. The *Locality (LO) layer* indicates from which or to which interface the signals are propagated.

5. The *Data-flow Selection (DS) layer* accumulates the relevant signals for a given processing unit.

6. The *Representation (RE) layer* constructs the input for the *CO layer*.

7. The *Cognitive (CO) layer* contains the actual processing functionality.

Chapter 3 discusses how the processes are actually going to be modeled. This will be done using network modeling techniques.

Chapter 4 shortly shows what has been done regarding literature research what hasn't been named (properly) in the rest of the thesis. This not only supports my goal, but also my approach. In this chapter a new scientific process is also proposed. After the process has finished the proposed process may serve to analyse the instantiated process itself. However, there isn't a set scientific method which may be assumed to be correct beforehand and thus applied as a procedural approach. The process may only be analysed afterwards.

Chapter 5 discusses the constitution of each of the layers in smaller processes linking them together. The signals as passed from the PW to a being are called *dormants*. These are passed via *receptors*. The signals generated by a being are called *actuators*. These are passed to the PW using *emitters*. The *receptors* and *emitters* form the interfaces with the PW. The internal communication is established by passing *transmittees* over the network. For each of the processing units the *transmittees* of a certain time-frame are gathered in *information* instances for each of the separate *receptors*. Together with the previous output of the CO process, named *desires*, a new input is constructed for the CO process in the form of the *potential wave*. This process then generates a new output and adapts its own process using the given input. The output isn't just used as new input, but also induces a process to the PW. This process then serves to generate *actuators*. These then can influence the PW as well as perception. They can possibly also be perceived themselves if appropriate *receptors* exist.

Chapter 6 describes the properties of the basic functionality which may lead to emotions in the form of an *Emotional Reference Model (ERM)*. This describes on the one hand the emotional content as it is represented within the *CRM*. The emotional content is captured within the physical properties of the being itself in relation to its cognitive process and is thus represented by the *potential wave*. After having established this, this chapter also describes the basic behaviour of the *CRM* using differential equations. This also explicitly defines the presence of a relation between beings and the PW. Just as well this clearly indicates that in case of beings we aren't dealing with simple *on* and *off* buttons, but with variable *real* input values. From these values then emerge certain expressions based on highly complex compositions of processes.

The uttered emotion now may be directly related to the emotion itself. This is not only induced by the emotion, but the emotional expression itself also induces the emotion. This can of course all be put in relation to previous emotion models which generally fail with the classification of *Basic Emotional States (BES)* from which *Composite Emotional States (CES)* are constructed. The CRM doesn't classify. The emotions express themselves according to a *fuzzy model*. The *CRM* contains the functionality of such a *fuzzy model*. The reflexes which are expressed in regards of emotions can also be modeled using the *CRM*.

In conclusion of the chapter certain mechanisms are proposed which are necessary to come to emotions. A being needs to be able to (learn to) identify emotions. A being needs to be able to (learn to) utter emotions.

Chapter 7 considers the foundation of the theory. This shows that it's not possible to ignore one of the layers. Just as well it isn't possible to change the order of the layers. Regarding completeness of the model no possible guarantee can be made that there isn't another layer, but until this layer is properly identified this model may be assumed complete. The likelihood that another layer is found, based on our estimation of how everything works (our "knowledge"), is very low.

Conclusions and Recommendation close off the theory. In conclusion it may be said that the goals have been met. It is however still very important to think about the ethical consequences. From these ethical consequences then may be discussed in how far architectures should be a simplification of consciousness. Such a simplification than means that the instantiation of such an architecture doesn't cross any ethical boundaries related to consciousness.