Derivative Ideas and Considerations Based on the General Theory of Consciousness

(Master's Thesis Appendix II)

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

Foreword

This very small paper was written at first as part of the larger paper on the *General Theory of Consciousness (GToC)*.[1] In order to separate essentials from non-essentials the original thesis has now been divided in three separate reports:

1. the GToC
2. Appendix I on agents[2]
3. Appendix II containing derivations from the GToC

This is appendix II.

[1] (Hobo, 2004b)
[2] (Hobo, 2004a)

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

Proposition

Definitions

Statements

Chapter 1

Introduction

What good is understanding consciousness and its mechanisms? One of the good things of knowing the basics is that now solutions and ideas found in consciousness may be derived in order to apply them in new ways. Some of these will adhere more closely to the actual mechanisms of consciousness than others and the types of solutions or ideas may also differ greatly. Based on the *General Theory of Consciousness (GToC)*[1] this part of the thesis discusses some of the derivations and considerations which may be associated with the GToC.

The derivations and considerations have been split into four parts. The first two are typical for the classical approach to AI. They describe *derivative* ideas from the GToC. The last two are directly related to the GToC.

First chapter 2 describes representative models for architectures which need to express feelings. Then chapter 3 describes different classical learning mechanisms. Chapter 4 discusses different philosophical. Chapter 5 discusses different solutions which have been found in nature and can now be described by applying the Consciousness Reference Model (CRM)[2]. Finally the conclusions review the possibilities of the GToC.

What should be noted is that this part of the thesis doesn't discuss a conscious being. It just discusses some ideas and solutions related to and found by consciousness.

[1] (Hobo, 2004b)
[2] (Hobo, 2004b)

Chapter 2

Feelings

In the past many issues regarding feelings have been used within AI. Having the GToC it now becomes interesting to look at the relation between what has been done and what should be done regarding AI and feelings.

I would first like to propose an example of a simplified derivative emotional model. This is described in section 2.1. The emotional model will serve as an initiation point for some of the discussions that may lead to architectures of conscious beings.

Strongly related to the theories of emotions are of course theories of humour. All the theories that are known come into expression in for instance human beings. Section 2.2 shows how different theories can be associated with the CRM.

In order to truly enjoy humour as we do, certain qualia need to be associated with the experience. Is it possible to describe these qualia and are they a necessity for the behavioural properties of consciousness? Section 2.3 explains the relation of qualia with the CRM.

Section 2.4 expresses the relation between experiences and how they come into being by describing the experiential loop.

2.1 An Emotional Agent-Architecture

In this section I'd like to propose an Emotional Agent-Architecture (EAA). This EAA should be able to cope with other agents within a certain limited world by expressing emotions and performing actions. These other agents could also express emotions and must at least perform actions within that world. To some extent the alteration of the world and the world itself may be seen as some sort of an agent. Anything that isn't the EAA itself belongs to the PW from the EAA's view.

The actions will consist of for instance the physical communication. Here each of the agents may influence each other and the world explicitly using those actions. Such actions may then be formed by for instance querying a database. Just as well they may be used for physical actions within a world. The actions are thus related to the make-up of the world within which the agents live.

The emotions will be needed to influence each other implicitly, maybe persuading each other to a different choice of actions. The EAA as proposed can

be suited to work in different types of environments. One could for instance propose a teacher-pupil relationship. Something entirely different may be a virtual world where the beings walk around and interact with each other as though naturally. This way a user may again enter that world where the agents have learned to mimic the right emotions given certain reactions from the users.

To describe this EAA there are two perspectives that will need to be dealt with. On the one hand there's the CRM. On the other there are the models that were already previously known to us and have been more or less proven to provide some seemingly natural interaction mechanisms.

First let's take a look at the CRM. In deciding which symbol or which value to utter you'll need to have a simple representative of a CO process. This symbol may then be represented by a contraction of a muscle mimicking an emotion or executing an action not towards but within the PW. These values may then be rewarded.

The decisions made should make sure that the agent performs as well as possible in its world. From this follows that the CO process should decide based on three values:

1. Its own current state,
2. the state of the PW and
3. the estimated induced rewards of each of the possible actions.

These values should then be used to control the agent in such a way that its behaviour seems quite natural to us.

To come to a more natural expression of emotion it's best to let the EAA learn to smile. This should provide emotional expressions that are more coherent than some in the models as we know them. This learning will have to come forth from positive rewards associated with positive emotions and negative ones with negative emotions. Another type of reward that should also be considered is the reward from the PW. This should indicate whether the action that was performed or maybe the expression that was uttered was desired or not.

The action, contrary to the expression, is not fed back to the process because only the state of the PW should influence the actions. The action in case of this agent isn't part of the PW nor is it part of the CO process. It is an induced change in the PW. The process which conforms to the action is modeled within the EAA. The expression and action values are a very simple form of actuators. These then induce a certain process to change the PW. Basically part of the PW layer and the processes which change relations within the PW are now part of the agent.

The way to picture the forming of expressions is by creating muscle groups or, more generally, expressional areas which utter rewards on stimulations. This way it should be able to learn to utilise these expressions more appropriately. The correct expression should then come forth from the highest expected reward in the future. This may also mean first accepting a negative reward temporarily which may lead to a higher positive reward in the future. The groups of muscles which are stimulated will have to pass through some kind of selection mechanism where for each muscle the desire to stimulate that muscle in a certain way is uttered and chosen based on natural selection. The way this is then represented on the face should be done by a natural representation of muscles. This may

also mean that when opposite desires are uttered a muscle may start to shiver. This will mostly be related to reflexive mechanisms.

The actions together with the expressions will influence the outside world. The outside world here is the world outside of the agent's body. The expression is part of the internal body of the agent, but is visible to the outside world. An action directly influences the outside world by changing it physically. The outside world will then process it's perception of the expression as well as the change commenced by the action. This change will of course be represented by the new state. The direct influence of an action on the PW will also be noticeable through the input of sensory data (for instance pressure). Here the action is perceived without passing the action itself.

The CO process will need to be made up of some kind of reinforcement learning mechanism. The size and dynamics of the mechanism shouldn't be confined to too limited boundaries. They should be large enough to ensure that the sought for behaviour actually arises. The behaviour we or any being expresses in emotions is very complex. This means that when the CO process is over-simplified and can only contain a too limited number of reference states or dynamical rules, this will leave us with a too statical architecture. Of course to ensure that the basic learning mechanisms work it is useful to first test with only very few very crude representations of emotions. But when the basic learning mechanism is ensured to work, it should be allowed to dynamically grow and resize to fit its needs.

Next to the CO processes it might be an idea to implement reflexes which weaken partially or totally when the CO process becomes more advanced. This way the CO process now has to compete with the reflexes. It is now possible to give the agent a knowledge-base to start from on which it may improve. The architecture in a graphical representation will now look like the one displayed in figure 2.1. The circles mean that signals are combined into one new input.

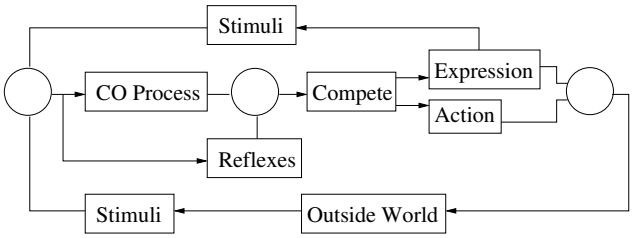

Figure 2.1: An emotional agent architecture

The way this agent is now modeled it can interact with a user using physical actions and it may also support its response by an expression. Research areas which may be interested in this architecture might for instance be concerned with tutoring systems. This way a tutor may learn to become angry or maybe even control its anger in some cases. Another example could be a training program for first-aid assistance. The reactions and actions of different types of agents in a world could be modeled using the EAA.

What now becomes interesting to look at is what kind of models have been proposed in the past. The GToC is very clear in the way it works and should be able to be related to other models. This is of course under the assumption that the models work properly.

One such model that has been proposed in the past is the Emotion-based Architecture (EBA).[1] There are a few things which may be noted.

The EBA discerns the CO process and the perceptual process. It acknowledges that learning can take place within both types of processes. This is also the way it has been done within the GToC. Within the GToC receptors may be freed and blocked within the PW layer. This may be used to free and block incoming signals physically. The CO process also learns.

The EBA also discerns something like the potential wave and the CO desires. The same functionality as present within the potential wave may be represented by the *desirability vector* (DV). This DV holds values for different types of emotions and decisions. Based on the DV, the perceptual as well as the CO process then generate stimuli. The DV in turn is influenced by the perceptual as well as the CO process. The perceptual layer directly generates stimuli from the DV. When the agent is built the perceptual layer already contains part of the functionality of the mapping process. So it may be said to contain reflexes. The map is adaptive. The CO process creates CO images, which may influence the decisions. Most of the decisions are based on the perceptual layer. The agent learns by receiving and processing a *feedback stimulus* from the environment.

A different kind of model which is also quite often used is the fuzzy rule based model. Although the fuzzy system should provide somewhat fluid transitions from one emotional state into the other, it does contain explicit rules. The downside to this has been clearly illustrated in the GToC.[2]

2.2 Humour, Social Position, Tension and Release

Different factors have come forward in relation to the GToC that are of a more or less direct influence on the principle of humour. There's trying to achieve a certain guarantee of continuation of processes. This will mean establishing a socially stable position, which may be established in different ways. There's also the charge of different receptors, which may induce physical tension. This tension of course will need to be released. When the tension becomes too high this will limit us in our capabilities of continuing our processes. All these aspects have come forward in different models concerned with humour. Humour, as you may well understand, comes forth from consciousness itself. What would be a correct way to describe humour in relation to the proposed GToC? Models which are evidently related to the models which will be discussed further on I leave to your pleasure to explain. I don't believe in one theory of humour, for there are many types of humour suiting different types of *purposes*. This section will try to provide an overview of what has been considered in previous humour research as well as relate it to the GToC.

To start with there are three conventional humour theories.[3] These are the Superiority Theory, the Relief Theory and the Incongruity Theory. Each of the theories may in some way be related to the other two. In the remainder of the section follows a discussion of these three theories, which may also serve to illustrate the lack of possibility of finding one specific theory of humour.

[1] (Custódio et al.)
[2] (Hobo, 2004b)
[3] (Mulder and Nijholt)

What does the Superiority Theory comprehend? It's based on the continually competitive relations between beings to be the strongest.[4] The competition should make sure that the process of a species continues. How does this continuation come into being?

There are three factors concerned with the Superiority Theory according to Gruner:[5]

1. *Every humorous situation has a winner and a loser.* The winner is basically allowed to continue its thought process and thus will determine the way the world continues.

2. *Incongruity is always present in a humorous situation.* Incongruity partially means that somewhere something fails. Identifying failing will also set a course where the failure will be prevented. The failure which could have caused the cessation of the process thus doesn't come into being and the process continues. *Failing* here means that a certain association or associative process cannot be established.

3. *Humour requires an element of surprise.* The element of surprise comes forth from wittiness, where the one outwits the other. The attack that comes as a surprise has the highest probability of success because of the attacked not being prepared to counter. Again this will make sure the superior being has the highest probability of continuing its process.

The Relief Theory stimulates certain parts of our physical processes to release the tension that's built up inside of us. This will mean that the tension is released by lowering the resistance and that a free flow of signals is temporarily allowed releasing the potential that was caught in a being's physique. This will thus also lead to the freeing up of the thought process. When the thought process is freed up and a being can think clear again about its position and possible actions, this will ensure or at least contribute to its survival.

The Incongruity Theory focuses on the change of the structure of the CO process. In case of humour the experience of the funniness of a joke is thus clearly correlated with the hearing and not previously knowing of the joke. In case of a joke that was heard before, it changes the CO structure in a lesser extent. It may also become so obvious that it doesn't have a positive reward associated with it but a negative one. But the experience of a joke as you may now see directly comes forth from the actual definition of experience itself.[6]

So in short the Superiority Theory focuses upon the continuation of being.[7] The Relief Theory primarily focuses on the release of tension or potential by creating actuators.[8] This also induces the continuation of being. The Incongruity Theory is directly related to experience.[9]

There isn't one specific correct instance of a humour theory. They are all more or less related and humour is formed in the same way as emotions are. There aren't any true classes of humour. If there were one correct theory this would be again a fully reflexive system. If this were so evolution would have

[4](Barnes, 1992)
[5](Gruner, 1997)
[6](Hobo, 2004b, definition 6.6)
[7](Hobo, 2004b, definition 6.14)
[8](Hobo, 2004b, definition 5.25)
[9](Hobo, 2004b, definition 6.6)

granted us all the same sense of humour. We know this isn't true. Humour to a very large extent depends on previous experience, present circumstances and different instantiations of being.

2.3 Qualia or Representational Feeling

One of the greatest mysteries of consciousness as identified by for instance Chalmers [10] is what makes us experience seeing *red* or anything else. As we long have argued from a philosophical point of view, everybody experiences an event in everybody's own way. Although we assume this to be true, this cannot just be accepted as truth, because we don't have physical evidence. But if we assume this to be true, this also means that we ourselves produce our own experiences.

Provided here is logical proof that we produce our own experiences. This will then also implicitly mean that every being experiences things in its own way.[11]

Just as well the failure of the requirement of qualia to exhibit conscious-like behaviour is proposed here. So what gives us this extra bit of experience is just a nice feature. An abstract model of this feature not truly containing this feature at all should be able to reproduce the same effect. The existence of this feature as we (claim to) perceive it is not a necessary given fact.

It isn't possible at the moment to provide anyone with an answer whether or not our PW representation of feeling as we know it and experience it is an abstract representation in itself. This is left to future clever scientists to prove or disprove. Here is just shown that the GToC is sufficient, where qualia may be an implementation leading to certain functionality, but not a necessity for the functionality itself.

First let's argue that we ourselves generate the feeling which we feel. What can be provided as an argument to show that we not only experience qualia, but actually generate them ourselves as well? Does the light not fall into our eyes? Isn't the ice-cream truly cold and does it not contain these qualities itself. Ah, here lies the difference! The physical qualities which we perceive they do contain, but not the qualia. To be able to show this it's best to look at deviations of the system which generates these qualia.

A deviation which is rather well known is synaesthesia.[12] This basically means the cross-linkage of different senses. When a sound is uttered, in case of some synaesthesists this may lead to the experience of certain colours. A pulse of light is a completely different kind of wave than for instance sound. The only way these colours may be experienced is if we generate the perception ourselves.

Another example that you yourself may experience is for instance when you (carefully) apply pressure to your eyes with your eyelids closed. Especially in the dark this will result in the experience of clear spots of colours. Funnily this may also work the other way around. When a blind person is given sight by surgery the perception of colours is often associated with pressure applied to

[10] (Chalmers, 1996)

[11] The statement that we experience in our own way isn't proven itself and thus cannot serve as proof.

[12] (The International Synaesthesia Association, 2004)

their eyes. This was because the only sensation of colours they had previously was by applying slight pressure to their eyes.

These two examples clearly illustrate that we ourselves generate our own perceptions. Considering the facts it's not possible to have any further useful debate regarding this subject. Now let's argue that true qualia are or could be nice if they existed, but they aren't a necessity. What do these qualia mean? What are they?

Qualia are the felt experience by beings. This isn't the same experience as in the experience gained from emotion[13]. The latter is the experience a being gains during its life-time so it can more easily respond correctly with the right actions in later stages of its life. The former so-called experience which we are talking about now is feeling. This feeling which we perceive is greatly influenced by our own internal state. For instance when we are really mad we often do not hear anything and start seeing black spots (or, to put it differently, our vision becomes blurred). This means it blocks our senses, because the resistance becomes too high for any dormant to get through. Suppose now there were some system, which based on our internal state, our physical internal state and our desires, influences our perception?

By emitting actuators, which may block and free different senses, the same behaviour may be implicitly reproduced which leads to for instance *angry* behaviour. Or maybe it will lead to a release of tension which may for instance be uttered in something like laughter. There is no need to have qualia as long as the mechanisms portray the same external functionality.

The final argument which is actually the most compelling is given by me in a paper about agents when discussing the senses.[14] It describes the reason to assume that qualia are mathematically induced. Basically it proves that the felt experience changes with experience gained during a life-time.

definition 2.1 (qualitative behaviour) *Qualitative behaviour is the behaviour portrayed based on the perceptual process as well as the cognitive process.*

definition 2.2 (perceptual process) *The perceptual process is a representative of the felt experience which changes based on gained experience during a life-time.*

The perceptual process as should now be noted is formed by one or more instantiations of the CRM. This then generates the perception as actuators. These may then be received again by these or other instances of the CRM to process the perception further. These actuators portray the same functionality as the qualia as we from a procedural view logically perceive them do. So considering the qualia as inputs they have a functionally associated output.

Since qualia largely depend on experience gained during a lifetime I would like to capture qualia in the following manner. Let's start to call the experience gained during a lifetime *knowledgeable experience* and qualitative experience the *felt experience*. Knowledgeable experience initiates a changing of the CO process or basically the architecture of the CO process. This architecture also includes the physical medium on which this process is implemented. So the larger the knowledgeable experience gained, the larger the change in the CO

[13](Hobo, 2004b, definition 6.6)
[14](Hobo, 2004a)

process. This will then lead to on the one hand the CO process incorporating physical pieces into its process. On the other this will mean that parts of its former knowledgeable architecture will be excreted. What is done with this excrement?

The excrement as produced by the gained knowledgeable experience isn't directly related to the processing of the inputs. It's a side-effect. But this doesn't mean that the excrement will not hit something. This excrement may actually hit a receptor and the conscious processes may try to explain this excrement. The amount of excrement of course decreases when the CO process doesn't change that much. This may for instance explain that a joke heard again is less funny. In order to try to explain qualia I would hereby like to introduce the following theorem.

proposition 2.3 (qualia) *Qualitative experience is a created and found association by the cognitive processes of the excrement produced in the changing of the cognitive processes themselves based on gained experience.*

2.4 The Experiential Loop

Things that play an important role in conscious experience or the belief of having conscious experience are, as previously mentioned, emotions. These emotions tend to swell up inside of us, sometimes taking over our rational thought. In my thesis on the GToC a discussion was held regarding emotions, but if we place this in an experiential context, how would we then describe this?

To discuss emotions seen from an experiential context emotions should be divided into three categories.

First of all there are emotions that were previously known to us and thus can be identified. When an emotion is clearly known to us and we can identify it, this means that we associate certain actions with it. These actions may lead to us dealing with the emotion and possibly influencing it. For instance when we are angry we may think of the action *breathe*. This may then calm us down.

Second there are emotions that we haven't previously experienced and need to be interpreted. In order to learn to deal with these emotions the circumstances should be clearly identifiable. Based on the circumstances or by trying and experimenting new actions may be associated with those circumstances. These may then serve to control the emotions or maybe change the circumstances.

A third category would be emotions that need to be reinterpreted. It may also happen that emotions have been identified wrongly and need to be reinterpreted. This means that new actions will have to be associated with the found circumstances. In order to do so based on the circumstances or by trying and experimenting new actions may be associated with those circumstances. Just as with emotions which haven't been experiences before, the actions may then serve to control the emotions or maybe change the circumstances.

Now let's suppose that we do not know the emotion or have identified it wrongly. At first we will most probably not deal with it properly. The emotion will thus remain and will induce processes which will enhance the emotion (i.e. actuators[15]). This normally makes the emotion more clearly identifiable.

[15](Hobo, 2004b, definition 5.25)

Although it serves a positive cause it may get out of hand.

The emotions may in time, when not controlled properly, when the tension or potential isn't released in some way, take over the being. So beings that are less apt to establish an idea of what their emotional state is have a larger probability of ending up in an extreme emotional state, possibly not being able to snap out of it anymore. Not being able to deal with emotions would suggest that the emotional interpretational mechanism is lacking in some respects.

What comes forth from this is the *experiential loop*. In experiencing a certain emotion this changes a being's being itself. It learns, acquires knowledge of life,[16] so to speak. Some people would call this aging, which I think is improper, because there are plenty of people who have experienced a lot but still behave like little children.

The experiential loop means experiencing a certain emotion and the enhancement of this emotion by the experience itself. To snap out of an emotion this on the one hand takes the recognition of the emotion, which will need certain strength of will and clearness of thought. Only when the will or desire[17] to break out of the experiential loop is strong enough then this may be established. Strength of will here means that the processes controlling emotions, keeping them under control, function properly and in unison. Clearness of thought may be associated with that certain receptors aren't blocked, which leaves these processes free to form a plan (an inductive association as proposed in section 3.4) to control these emotions.

[16](Hobo, 2004b, definition 6.6)

[17]This *desire* may now be taken two ways: as the felt desire or the desire as uttered by the CO process (Hobo, 2004b, definition 5.22).

Chapter 3

Learning

One of the important aspects of being is growth and maturation. During growth and maturation beings learn and grow accustomed to their new form and the place they occupy within certain social structures. But how do beings learn, seen from the perspective of the GToC?

Positive and negative stimuli are presented to us based on which we learn, but why are these stimuli positive and negative? How did they come forth from evolution and are they then part of the CO process? Section 3.1 deals with appraisal theory.

To actually implement a form of reinforcement learning section 3.2 proposes a new model, namely the categorised reinforcement learning model. The model is loosely based on the principle of the potential wave.

In order to deal with induced processes there are different ways to remember the former input state. In case of reinforcement learning this is explicitly modeled in the relations between actions and states. In beings this usually isn't. Section 3.3 describes a method to shift time-frames using the fading of previous inputs.

In case of consciousness the models are not only less statical in make-up, but also in behaviour. The main idea behind the subsequent stages in establishing relations as well as reasoning is considered in section 3.4.

3.1 On Appraisal Theory

When a being learns, one of the mechanisms it uses to learn is reinforcement learning. In reinforcement learning rewards are uttered based on the actions chosen. These rewards may be positive or negative. The eventual goal is to reach the highest possible average reward. But what makes a reward high and what makes it low? How do these rewards come forth from evolution? Should we let these reward-mechanisms evolve or should we implement them ourselves?

Evolution[1] is the collection of continuations and cessations of processes. A positive reward is meant to induce the continuation of a process. A negative reward is meant to stop a certain line of processes or, to be more precise, to divert it and change to another direction in the kind of processes. This leads to definition 3.1.

[1] (Hobo, 2004b, definition 6.14)

definition 3.1 (reward) *A reward is a stimulus which makes sure a certain (cognitive) desire*[2] *is either strengthened or weakened.*

This may have two different kinds of influences on the evolution of species. On the one hand there are more *primal* properties of beings and on the other more *socially sophisticated* properties.

Seen from a world perspective it leads to implicit laws like *survival of the fittest*. With the way the food chain is established the basic idea behind survival is *eat* or *be eaten*. In order not to be eaten a being needs to be as fit as possible. This doesn't apply directly to our society though.

It may also be regarded from the internal view of being, which would indicate survival of the smartest. The smartest would here be the one with the fittest set of brains instead of the fittest set of muscles. Since they are laws, they quite often fail or need exceptions. That's why we can't really state them generally. Of course when looking at the way our society is organised, society quite often is reduced to *eat or be eaten* from an economical perspective. But this has been reduced to specific niches of society.

Whoever is the fittest depends to a large extent on present circumstances. Consider in nature a mouse fighting a lion. The most probable thing to happen is that the lion eats the mouse. But suppose the lion trips and breaks something, making sure he can't even chase the mouse: the mouse may outrun the lion and the lion will die slowly. By accident, the lion no longer is the fittest. This introduces fuzzy variables into the laws, which leads to them not being strict laws at all. The laws are for a large part subject to circumstances which can't be modeled completely. For a law to hold, it should contain no ambiguity regarding the interpretation of the law.

The actual reward mechanisms as we have them now can be quite easily explained as to why they came into being the way we know them. Suppose that a reward mechanism that leads to the failing of processes is instantiated. This will mean the process will fail. Since the reward mechanism is associated with this process, this reward mechanism will cease to exist as well. The reward mechanisms which ensure survival or *continuation of processes* then of course survive as well.

The reward mechanisms that lead to positive rewards regarding the continuation of processes have become the most frequently established reward mechanisms in nature. This also defines our sense of *true* and *false* or *right* and *wrong*.[3]

Of course there are even today sometimes emergent processes which contain negative reward mechanisms. This can be shown by the existence of for instance sadomasochist relationships in human society. But this isn't an exception because these mechanisms in the end will again lead to the cessation of processes. There's a relatively higher risk involved with the life of these people.

In case of consciousness what's now needed is a fuzzy reward system. This fuzzy reward system could theoretically be evolved. This would also mean the evolving of an entire being to start with. Introducing the mechanisms as proposed in the CRM, combined with evolutionary programming[4] this should be possible.

[2](Hobo, 2004b, definition 5.22)
[3]See definition 4.1
[4]See for instance Bongard's (2003) work.

Here the main goal would be the increasing of the lifespan of global and sub-processes. A more global process would be the survival of life and relative sub-processes then the survival of a species or a being.

The thing just is that evolving such a reward mechanism and different kinds of species will take ages. Of course it could be possible to create a simplified world where beings don't really need complex motor systems, but just move when they want to move. These kinds of worlds do have some theoretical value and may be used initially. In the end however we are more interested in our own world.

Instead of creating a simplified world it would also be cheaper and faster to implement a reward system yourself. This reward system will have to be approximately the same as implemented in us. This way it has become the same as the reward mechanism that has evolved in our species. If you let it evolve further from there, it will behave according to evolution. Such a reward system may then be used within an agent-architecture, because it has been proven to work within us.

What is now needed is the rewarding of desires associated with the actual choice of action. When your eyes are burning it doesn't help to throw water onto your hands (assuming they aren't burning also). This means that each of the senses will basically be coupled with its own reward mechanism. Here it should also be clear that there is a difference between each of them. All the rewards should now be presented as an input to the CO system. The CO system should in some way associate them with certain states. Based on these states it may then choose the appropriate action.

By rewarding a certain set of desires, other desires which aren't rewarded will still have a relative change in reward. The absolute estimate reward of these other desires will however not change. This will mean that if one desire strengthens, the others look relatively weaker. What will this now mean for the CO process?

Within the CO process two basic processes are of concern regarding the stimuli which are received. On the one hand there's the reward mechanism. On the other there's the actual generation of the output desire. The stimuli should now not only be coupled with the generation of the output desire, but also with the reward mechanisms. This way the stimuli may serve to determine the current state of affairs as well as be used to change the actual CO process.

In a conscious being the being may learn how to ignore rewards by basically creating different stronger desires and inducing a low receptiveness to the rewards by emitting the right actuators. The potential that does come through at the CO process should be directly coupled to the reward mechanism. This may mean that a being may learn to ignore for instance its negative stimuli in order to make the average reward higher.

statement 3.2 (expectancy of rewards) *The process of consciousness may lead to the ignoring of reward mechanisms in order to receive a higher reward in the future.*

This will of course mean that beings will start taking risks in order to save the species or just to obtain a *rush*, so to speak. This is also illustrated by the presence of for instance adrenaline in our bodies. This today quite often stimulates risky behaviour, even though it was meant to save us from dangerous

situations when for instance chased by a predator. Sometimes this also leads to examples of evolution in for instance high-risk accidents with a sport like base jumping. It may of course also still serve the purpose of attracting mates by showing strength.[5]

3.2 Categorised Reinforcement Learning

In the CRM the potential wave is an approximation of, or an infinite sequence of, potentials. It contains one potential for every receptor. This means that receptors may be categorised. With a small jump in thought this gave me an idea.

Suppose that someone was to receive a large number of small penalties within different penalty-categories. This will seldom feel as hard as a large penalty in one particular category. It often amounts to strain, but it will not necessarily amount to pain. This happens just as well with positive rewards. Note that I here don't mean that receiving a million separate dollars will feel any different from a million dollars in a single shipment. This all falls in the same category of money. Different categories may be a pinch in your shoulder, arms and legs. All these together will not feel as bad as a cut in your shoulder. The cut will bleed.

So how can it be possible to compare the categorised types of rewards? If you now have the different values for the categories all stored in one state, the total reward can be calculated by adding an exponential function of all the rewards together. This way one large reward will come out on top relative to lots of smaller rewards. By storing the rewards per category it's also easy to maintain the memory function per category.

The value of a certain state $V(state)$ can be updated where the old value will be partially remembered.[6] To do this a certain step size α will need to be taken into account where the step size will differ between *0* and *1*. The update rule looks like equation 3.1. Here s is the new state and s' the previous state.

$$V(s) \leftarrow V(s) + \alpha[V(s') - V(s)] \tag{3.1}$$

Equation 3.1 will now have to be generalised for a multi-category reward, with multiple values $V_{category}(s)$. Of course the step size may now also differ per category because some things are remembered better than other things. The update rule will now become like in equation 3.2.

$$\forall c \in categories : V_c(s) \leftarrow V_c(s) + \alpha_c[V_c(s') - V_c(s)] \tag{3.2}$$

To calculate the total reward for a state $V(s)$ we now need to add together exponential functions of all the rewards in all the subcategories $V_c(s)$. This is done in equation 3.3. Some rewards may influence the actual rewards more than others, so I've chosen to also let the exponential function (or factor) depend on the particular category. Of course the fact that some rewards may be negative still had to be properly modeled in the summation function.

[5] This final statement comes from Auke Pols.

[6] (Sutton and Barto, 1998)

$$V(s) = \sum_{c \in categories} \begin{cases} 0 & \text{if } V_c(s) = 0; \\ \frac{V_c(s)}{|V_c(s)|}(e_c^{|V_c(s)|} - 1) & \text{if } V_c(s) \neq 0. \end{cases} \tag{3.3}$$

Now it's again possible to choose the action which has the highest probable reward. This will be done according to equation 3.4.

$$a \in action\_set(s') \mid V(s(s', a)) = \max \{V(s(s', a')) \mid a' \in action\_set(s')\} \tag{3.4}$$

3.3 Fading Input Learning

This section describes the fading of input values as proposed earlier regarding the shifting of time-frames. This can easily be related to other work in different fields. An example is part of the Optical Flow Algorithm by Horn and Schunk. In a demonstration of this algorithm I witnessed that previously taken picture-frames were partially merged with the new picture frame. This same mechanism should apply for receptors during an infinitely small time-frame.

Suppose a receptor at time-step t_k receives a certain input value i_{t_k}. Now for every subsequent time-step t_{k+1} equation 3.5 should hold. Here α represents the replacement constant in order to indicate to what extent the old input should be replaced by the new actual input value $i'_{t_{k+1}}$. As may be expected α takes a value between *0* and *1*.

$$i_{t_{k+1}} = \alpha \times i'_{t_{k+1}} + (1 - \alpha) \times i_{t_k} \tag{3.5}$$

The step-size t_{step} for each time-step should be infinitely small. Mathematically this can be represented by equation 3.6.

$$\lim_{t_{step} \downarrow 0} t_{k+1} = t_k + t_{step} \tag{3.6}$$

For a bottom-closed iterative system the only thing that should now still be stated is that for the initial moment in time t_0 equation 3.7 holds.

$$i_{t_0} = i'_{t_0} \tag{3.7}$$

Because most systems have multiple input-values, i_t should be treated as an input-vector. The minimum length $length(i_t)$ of such a vector is then one element. The former scalar α becomes a $(length(i_t) \times length(i_t))$-matrix.

3.4 Reason and Association

Within classical artificial intelligence a learning mechanism is trained and then basically forms an estimate function with one specific output for every specific thinkable input. This quite clearly is a shortcoming in the architecture, because now there is only one possible answer which may be considered. In reasoning however multiple associations may be drawn after which a being determines which is best. This section proposes a basic model which describes the process of associating and reasoning.

Different aspects should be considered when establishing whether a relation holds. First of all, one relation's validity doesn't always necessarily dismiss the other's. In establishing what to see this largely depends on the estimated reward in the future.

In case of game theory, the best action is the action which under given certain circumstances would obtain the highest reward. The estimating of a relation or association by reason corresponds to this same idea. But instead of learning one specific state-model like in reinforcement learning, a being should be able to make an estimate of the current state as well as the associated reward. This isn't an a priori given fact as in most instances of applied reinforcement learning. Within reinforcement learning the state-model becomes fixed and has values associated with it. Within consciousness a being reasons about the state as well as the corresponding value.

When reasoning about a certain relation a being should then be able to come to an agreement in thought. This agreement will have to be associated with the existence or non-existence of a best estimate of relations. This best estimate comes forth from narrowing down choice to a specific set of relations until one relation emerges as the most profitable estimate.

The process of estimating the validity of relations can now be established by three consequent steps. These are repeated in a recurrent manner until the final association with a relation is established. The three steps are:

1. **Association.** Before a likelihood-function of any kind can be established a best guess should first be made regarding what the relation is. This association will be based on learning from past experiences and obtaining rewards. The being will basically have to learn to fit perceived sensory images into relations, just like a baby which needs to learn to recognise the faces of its parents for instance. The more complex the task the larger number of years this may take of course.

2. **Sensibility factor.** When associations are drawn, they are drawn based on highest likelihood. This may mean that a predicate doesn't fit within a certain relation. The amount of sense a relation makes could be called the sensibility factor. The sensibility factor makes an estimate of the found association to be true or basically guesses the estimated reward associated with an association of relations.

3. **Equilibrium state.** The sensibility factor will have to be fed back to the current guess of relations itself as a reward mechanism. At a certain point in time the association will reach a certain equilibrium state where the meaning of the relation doesn't change anymore. This will need to be determined to form the output association.

In discussing emotions in terms of the GToC, it was already established that the CRM is a dynamically influenced model. Just as well the CO process was known to have certain properties as described in the GToC. On the one hand the above described steps describe properties of the CO process which should also be established, i.e. it should be a process which is constantly dynamical through time. On the other hand it also describes the dynamics of multiple instances of the CRM working together. The different relations are then made up by the changing of resistance of receptors as well as the CO processes. This

will then make sure that a collection of instances of the CRM may be discussed in terms of a single instance of a CRM again.

Chapter 4

Philosophy

Although the GToC is a highly abstract mechanism and seems to be very technical, it may be used to describe philosophical aspects of being as well. This section describes a few of these philosophical aspects.

Some of the terminology used within the GToC is to us very clearly associated with certain feelings. But the terminology shouldn't be mixed up with the feelings we know. The terminology is just a collection of abstractions used in the GToC. Therefore it may seem plausible to show some of the relations between the GToC's terminology and actual feelings. This is done in section 4.1.

A rather large problem to beings is finding the right answer. The main problem with finding the right answer is discovering the right question. How is this associated with the CRM? Section 4.2 establishes the relation with asking questions and the CRM.

Because everything is associated with establishing relations within conscious processes, it seems plausible to be curious regarding remembering relations. Section 4.3 discusses memory and associated principles.

A thing worth discussing is the question what would happen when signals didn't arrive anymore from our senses. What would our brains do? Section 4.4 proposes a hypothesis on the generation of a world image when no world is provided. The hypothesis should still be tested.

Finally section 4.5 discusses *absent-mindedness* in relation with the theory as proposed in this thesis. It proposes some basic theories which still need to be tested.

4.1 Beliefs, Desires, Intentions and Truth

As previously stated in the GToC the output of the CO process is generated as a so-called desire.[1] The generated desire will here be called a CO desire. Why are these CO desires not beliefs or intentions or truths or why not even felt desires? Or are they maybe, also? Is there a reason to start a discussion about this or is my terminology correct?

The CO desire will try to induce a certain PW and maybe implicitly CO process. The CO desire doesn't always succeed because of some other uttered

[1] (Hobo, 2004b, definition 5.22)

CO desire being stronger. That's why it's called a CO desire: it wishes to initiate a process but isn't always able to. How may this then generate felt beliefs, desires, intentions and truths as we perceive them? And are these truly beliefs, desires, intentions and truths or do they just fall into these categories to certain degrees after which we claim them to be what we call them?

We sometimes believe something to be true. To be able to discuss the belief that something is true we then first must have an idea of what truth really is. Definition 4.1 illustrates truth.

definition 4.1 (truth) *Truth is that a stated relation under stated circumstances not only is said to hold, but actually holds.*

Although we may presume that all things behave according to a finitely definable set of laws, we cannot state that we can predict everything for certain. We can only look down to a certain level of things to estimate present circumstances. Below a certain level it's impossible to look further down, because things just become too small for us to perceive. To some extent the state of these things may be measured. Unfortunately by measuring these things we introduce a very large influence forcing these things in a *stable state*. It isn't possible to measure a *wandering state*, where for instance a particle may wander between two different states. When we perform a measurement we force the particle into one of these two states. We can only make an estimate of which one of these states this will be.

Most unfortunately a being cannot see and know all. In order to estimate truth it needs to be clear what the chances are that something perceived to be true actually is. Definition 4.2 clearly illustrates how truth is estimated.

definition 4.2 (estimate amount of truth) *The estimate amount of truth is the probability that a certain stated relation under certain stated circumstances not only is said to hold, but actually holds.*

Truth can be deduced step by step. The opposite relation given the same circumstances will thus not hold and will thus be false. Do we know truth? No, not really. We are only capable of making estimates of truth, where in some cases the chances of the estimate to be correct are higher than in other cases. This is nicely illustrated by us being wrong for quite a large number of times. Here follows that a belief in turn is just an estimate. This estimate is made by our CO process to assess our current situation or state. The belief will be formulated within the PW by processes stimulated by actuators.

definition 4.3 (belief) *A belief is an estimate of the amount of truth.*

A desire to reach a certain state follows from certain beliefs. On the one hand there is the belief that a particular state is better or feels better than the current one. On the other hand there is the belief that the state can be reached by commencing in certain actions. These actions are caused by actuators which induce certain processes and these actuators have been induced by a certain CO output which we have formally named desires. Here we will refer to these as CO desires, not to confuse them with the feeling called desire. This links the perceptual desire to the CO desire to utter a certain actuator leading to a certain PW process.

definition 4.4 (feeling called desire) *The feeling called desire is the belief that certain actions need to be undertaken inducing certain changes in states that will enhance the chances of the continuation of processes.*

Just as well as the CO desire may generate certain actuators. These may stimulate actions which lead to a certain state. But sometimes we fail because of misjudgment of factors. Basically the intentions to do something serve the same purpose within the PW as the CO desire does in our conscious process. They try to create a certain state but the state is not always achievable due to external factors.

definition 4.5 (intentions) *Intentions are the instantiation of actuators to stimulate actions taken to reach a certain state where it may not be guaranteed that the actions are actually taken or the state will actually be reached.*

Although we commence into certain actions, different actions may be stronger and will thus often succeed. Sometimes through different circumstances the seemingly less strong actions may lead to the reward. This is the main thought behind natural selection. Usually it means survival of the fittest, but sometimes the less fit profits from certain circumstances. This of course means that it contains a different kind of less perceivable fitness in different areas. The fitness is then largely subject to circumstances. Unfortunately with the way how people interpret fitness, the term *fitness* usually leaves too much room for ambiguity. This is of course directly associated with *natural selection*.

definition 4.6 (natural selection) *Natural selection is the collection of processes which continue and thus not stop within a given set of relations.*

4.2 Asking the Wrong Questions

Some people always answer by question. Instead of providing an answer which contains insight, merely a door-way to insight is provided by posing an appropriate question. From this question then almost immediately should follow insight (basically the answer). Not only from a philosophical point of view, but also from a researcher's point of view, the problem usually isn't finding the right answer, but posing the right question. When the question is found the only things necessary to provide the answer is the language in which to provide the answer as well as the time and effort to write the answer down. Funnily this small piece of philosophy also provides us with an answer to how problems are solved in beings.

Suppose a being is looking for a cup. For a change the cup isn't empty, but contains coffee. The being knows it had the cup previously but has put it down somewhere and doesn't know where it is anymore. Where did he put *it*?

What is *it*? *It* without a context can be anything and is incorrectly stated for our brain to solve the problem in quite a lot of cases. In posing the question "Where did I put it?", a being is most likely to picture many places and look at those places without seeing the actual cup. To find the cup, the being should be focussing on the cup in combination with the place to find it, instead of only on the place. When only focussing on the place to find a general *it*, *it* will most probably be overlooked.

Suppose the being now pictured the cup and not for instance the cupboard. Suppose the being wasn't just looking around, but actively focussing on the memory of *the cup* or maybe *any cup*. Then the CO process would have a relation which it can attribute a certain value to. Because the CO process can determine a relation which may or may not hold, the chances are relatively larger that the cup is found.

So basically a being should generate a certain context, the question, into which the answer should be fitted. This frees up different paths (receptors of different smaller CO processes) within the larger CO process into which the relation fits and may be determined to hold. By just initiating on this path, through a process of induction the answer will be directly associated with.

statement 4.7 (iteration of processes) *The establishment of a context or plan comes about by lowering the resistance of receptors and thus freeing subsequent cognitive processes where each process is now easily stimulated to take place by each preceding process.*

An application of this principle can be found within sports.[2] Sports uses different kinds of visualisation to improve performance. Improving performance is done by fine tuning the mind. This can be done to obtain a right frame of mind, but also to visually rehearse techniques or maybe the whole of the game. For instance in cycling the participants often rehearse the tracks so they know every curve of the route. This makes sure that they can plan ahead and basically imprint when they should peddle to the fullest or maybe hit the brakes.

4.3 Memory

Back in the days when we were very young, everything was new. We quite often remember it like it were yesterday. What makes sure we remember it? What makes sure our minds wander off and sometimes revive these same sensations which we had when we were smaller? To do this we need some kind of physical representation of memory. How do we store ideas and perceptions? What do we actually store?

When a being has memory of something, this memory is bound to some LO. Because it's associated with a certain LO this means that it establishes a certain context within which this memory fits. To be able to establish this context this memory must have happened previously. Our brains do make mistakes because we sometimes believe some memory to be true even though it never really happened. This then usually comes forth from fantasy or, to put it differently, from contexts that were formed by partially self-generated input by our brains. There are different theories about why this happens, but this is beyond our present context.

statement 4.8 (memory) *A memory establishes a certain context.*

As statement 4.7 clearly describes a context is established based on subsequent CO processes. Each process here basically contains a small piece of the larger memory. Each piece establishes a small part of the larger relation where the total of all relations establishes the logical context. But in having a memory

[2](Ungerleider, 1996)

these relations must be established more permanently. On the one hand the processes are relatively permanently linked and form a network. On the other each CO structure has been trained in such a manner that it contains a very small piece of the relations which make up the memory.

What now happens is that when a being recalls a memory, the being will reestablish the context. This means the receptors are freed up lowering the resistance. After this the memory can be stimulated and the relations are reestablished. In establishing the relations either for the first time or reestablishing them again, the CO process processes these relations and utters desires. Based on these desires an image is built of the current scene or the scene a being remembers. This image is built within the PW by emitting actuators stimulating the processes which create the actual imagery or maybe the imagery is contained within the actuators themselves. These images may then be perceived and in case of a recollecting memory: perceived again. This then makes a being perceive or remember a certain event, basically establishing the context.

statement 4.9 (image generation of memories) *It doesn't matter whether a relation is established for the first time or not in establishing an image by generating qualia from the relation.*

The above describes the general idea of memory. There are however quite some features associated with memory. There's the amount of memory a being has, there's the vividness of the memory and the period of time a memory lasts.

The amount of memory is clearly related with on the one hand the size of the CO structure and on the other the dynamics of the CO structure. Suppose we would look at the CO structure as being something like a basic hard-disk, the amount of memory would be directly related to the number of smallest CO processes the CO structure contained. Each CO process would now be associated with for instance one value.

Instead of just being associated with one value, each CO process forms a functional representation. Here the representation is related to all the input values resulting in a certain output and even change in CO structure within the CO process. This way one CO process may be associated with multiple relations or (in this particular case) memories.

statement 4.10 (amount of memory) *The amount of memory comes forth from the size and dynamics of the cognitive structure.*

The vividness of the memory comes forth from how well-established a relation is. Although the vividness fades in time, using certain techniques like for instance hypnosis it's possible to reestablish the relations and strengthening the relations. This will then lead to a more vividly felt experience of the memory. So basically the vividness comes forth from the strength with which a certain relation holds.

The strength of a relation depends on the one hand on how easily the currently established relation is disrupted. In a non-stable system this means that a small push will disrupt the relation. In a stable system it will take a large push to disrupt the system. The extremity of the relation comes forth from the amount of which a relation can still change in both directions. If the amount a relation can change into one direction is null and it can change in the other direction, this is an extreme.

statement 4.11 (strength of relations) *The strength of a relation is established by the extremity and stability of the desire uttered by a CO process.*

statement 4.12 (vividness of memory) *The vividness of memory comes forth from the strength with which a certain relation holds.*

The period of time a memory lasts comes forth from the speed with which the vividness of memories fade. In case of short-term memory the relation will only be captured in the CO structure for a short period of time. This means that the relations established by the CO structure degrade rapidly, loose their structure, and thus do not hold any longer. In case of long-term memory the relations are more well-established and the structure is much more fixed than with short-term memory. This will mean that relations will hold much longer within the CO structure than they will with short-term memory.

statement 4.13 (lasting of memory) *The length of the period of time a memory lasts is reversely proportional to the speed with which the strength of a relation degrades.*

4.4 World Generation

One thing which has provided us with a lot of philosophical debate is the principle of dreams. Since we are not really commencing in PW activities and experiences they don't seem to be related to consciousness.[3] At least not in the way proposed. The most common theory of dreams is that they release tension and let us deal with things for which we didn't have any time to deal with during the day. But this does mean that our brains can generate images without visual, auditory, haptic or chemical stimuli. What would our brain do if these stimuli were absent as a whole?

The hypothesis here proposed is that the brain has certain functionality which does its work regardless of the presence of physical stimuli. When the brain is alive even though stimuli are absent it will keep on associating. This will then imply that it has to generate perceptions of stimuli itself even though there really aren't any stimuli. From a psychological point of view one could state that even though there is no PW to associate with, the brain wishes to associate and thus creates its own world.

In order to test this hypothesis a harmless experiment can be conducted with people who have lost one or more of their senses during their lifetime. People who haven't had any perceptual knowledge prior to their lack of perception can't be used because they don't have a reference frame to state that they have some kind of perception. Maybe also some best guesses could be made regarding people without any senses by measuring their brain-activity. Of course all proof is subject to a rather high cheating-factor in case of consciousness, so even proof without referencing such a person can be considered.

What now needs to be done with people who have lost one or more of their senses is to establish the following things. For each of the senses people should be asked whether they still experience some kind of perception of the specific sense. What should also be made clear is whether the perceptions are related

[3](Flanagan, 1997)

to the perceived perceptions in the past. The hypothesis is that they constitute new kinds of mental images. They may show some coherence with PW images shortly after the loss of senses, but when memory degrades they will become new types of images. Maybe these images will be mood-related, but this should only be tested later after the mental imagery has been acknowledged.

It may actually also be a good idea to scan for brain activity with any type of senseless person. I don't know what this will do. Functions of parts of the brain may also change when these parts aren't needed to deal with the specific sense anymore. So it isn't completely clear to me what brain activity will mean at such a moment, but maybe with different functionality the perceivable type of brain activity also changes. I don't know this.

4.5 Absent-Mindedness

Something which has been mentioned previously in this thesis which is an important aspect of conscious behaviour is absent-mindedness. This basically look like a wakeful unconscious state. This section illustrates three types of absent-mindedness which need to be explained as separate states. Before these three types of absent-mindedness may be discussed first some basic observations have to be made.

In basic every day processes we are most probably absent-minded more often than we actually realise. For us to notice that we are absent-minded there are a few assertions which need to hold. When we are only absent-minded for a split second no-one is going to notice, unless you pay direct attention to it. In order to notice that we are absent-minded, the following two assertions need to hold.

1. We need to be absent-minded for longer periods of time.

2. Our absent-mindedness needs to end abruptly.

As stated earlier there are three types of absent-mindedness. When both the above assertions have been proven to hold it's possible to identify what type of absent-mindedness actually occurred. What types of absent-mindedness are there then?

First of all there is absent-mindedness which may be related to active associative processes which don't take PW signals into account. Absent-mindedness here means that our entire process of thought is focussed on the internal thought process. This in turn means that there's no reference to PW processes within the internal thought process. The internal thought process may here be considered as the thought process which only considers memories and creating associations. What then brings forth this type of absent-mindedness?

What I would like to pose as a theory is that this absent-minded state is a state in which we are emerged when we need to find associations. When the associations are very simple, what we need to do is so obvious that this almost happens reflexively. But still we sometimes need to re-order our thoughts to solve basic tasks. When the association has been formed we return from this absent-minded state and execute the necessary actions according to the associative paths which have previously been formed.

Now suppose that we are in a complex situation which we need to think about in order to come to a solution. The more complex the situation the

more thought needs to be spent on the subject. This may lead to us living out different scenario's in our head in order to choose the right one. This then leads to longer periods of absent-mindedness.

The second absent-minded state which may be identified looks like an absent-minded state which doesn't refer to PW nor CO processes. Suppose that there are some processes which we don't even perceive which take up a lot of time? Suppose that some unconscious problem solving is going on which takes up all processing time? In our perception of the mind this will look to us like our though-process was entirely empty. So we have the idea that we weren't having any idea at all.

The third absent-minded state may be identified as absent-mindedness as found in meditation. Of course people may actively or without a previous plan enter into a meditative state. This is a different kind of absent-mindedness, since it acknowledges the PW signals, but just doesn't do anything with it. So the associative processes don't do anything. Instead the associative processes are allowed to relax. This may only happen when these don't receive any inputs any more, i.e. they are fully blocked.

These three types of absent-mindedness are three very distinct kinds of absent-mindedness. The first one only considered CO processes, the second considered nor CO processes nor PW processes and the last one only considered PW processes. When in search for what physically brings about absent-mindedness, people need to realise which of these three options they are actually looking for.

The way they are described here each of the processes may be described directly by the GToC. In order to do so, there will need to be a clear classification of processes. Just as well it needs to be clear whether qualia are built for the CO processes and the PW processes. Based on such a classification the GToC then may be directly applied.

To test the above theories people could voluntarily be scanned regarding brain activities over longer periods of time. This may then lead to the acknowledgement of one or more of these kinds of absent-mindedness. Possibly a further subdivision needs to be made in absent-minded states. The three described states are sufficient as top-level types of absent-mindedness though.

Of course in testing people they shouldn't be told that they are being measured for absent-mindedness, because they would try to control their thoughts too much. Instead they would have to be monitored for over a week with an observer noting their activities and inactivities. People could be asked to perform simple problem solving tasks to separate the thinking processes from activity processes. Of course it wouldn't hurt measuring a Zen-master during daily activities, because people like this have a highly trained mind. It's well worth the effort to also measure people with extremities in thought processes. Because of these extremities it becomes much more clear what patterns are associated with what behaviour. Participants should all be monitored in daily life as well as during specific small tasks where they should all receive the same tasks.

Chapter 5

Solutions

There have been some classical problems in the past regarding perception or even information filtering which have or haven't been solved. It is of course interesting to discuss these problems in the light of the GToC.

One particular problem that has previously been described is *the Cocktail Party Effect*.[1] I would now like to discuss how the solution to this problem may be produced using the GToC in section 5.1.

5.1 The Cocktail Party Effect

I'm not a huge fan of cocktail parties, but I have been present at many parties in the past. I will most probably visit parties in the future as well, having the same old problem over and over again: how to understand your neighbours which you're speaking to? This effect was dubbed by Handel[2] to be called *the Cocktail Party Effect*.

Within research areas involved with human-media interaction and particularly speech processing, partial solutions have been found to this problem. Different mechanisms come into play in solving the problem. Of course these mechanisms aren't perfect. We ourselves often aren't capable of filtering out other people's words. In case of a problem like this there exists no perfect solution, just best estimates.

If a being needs to focus on a certain signal, the being will have to make an estimate of the highest likelihood as to what that signal might be. This highest likelihood may depend on a few different things:[3]

1. Location of the sound source,
2. difference between the reception of sound by spatially separated receivers of sound and
3. separation of sound into different frequency ranges.

The likelihood may be enhanced by visual aid, but this will just help in determining the location. Although it will help it's not a necessary aid. For instance

[1] (Hobo, 2004b, section 5.6).
[2] (Handel, 1989)
[3] (Arons, 1992)

blind people manage quite well on hearing alone and don't need visual help. Because of this it's now possible to ignore the influence of sight in this discussion. How do these issues physically come into being?

The location or at least direction of the sound source is determined by phase-shifts of a certain sound between multiple receivers of sound. A practical example of the use of phase-shifts is found in air-traffic control.[4] By creating a phase-shift between the two speakers of a headphone for the air-traffic controller's voice, it became a whole lot easier to understand what he was saying. Normally it would be really hard because of the large amounts of noise made by the aeroplane. By shifting the phase it's easier for human beings to find the virtual location or *source* of the voice and thus the voice itself. But this is just a practical example of making use of the mechanisms in our own being. How can we recreate these mechanisms and utilise them appropriately?

In order to deal with recognition of sound better using the technology we have today, sound is separated into different frequency ranges. In case of our ears, this is also so.[5] Looking at the make-up of our ear it can be easily perceived how this is done. When a certain sound hits our ears this sound will pound the ear-drums. This will make sure each ear-drum starts to vibrate with certain frequencies, not unlike the way a microphone is stimulated. In case of a microphone however this is directly translated into a certain potential. In case of the ears it works quite differently. Within the ears the ear-drum passes its vibration through the middle ear-bones. These bones then let the fluid in the cochlea vibrate with this frequency. Within the cochlea there are spread along the rolled-up length of the cochlea lot's of minute little hairs. Each hair is basically associated with a frequency of its own. When this frequency is contained within the sound the hair will start to vibrate and generates a signal which is passed through the auditory nerve to the brains.

In case of a microphone an electrical potential is generated. The potential that's generated is the representative of the pressure applied by the ear-drum and the middle-ear bones on the fluid in the cochlea. Since it's just the pressure at a certain moment in time, by passing it directly to the CO process this would then pose a large problem. The CO process would have to establish a way to decompose the sound into different frequencies itself over a certain time-period.

So basically the signal from a microphone still has to be decomposed into a very broad range of frequencies before passing it to the CO process. Each of the stimulations provided for each of the frequencies may then be provided as an input signal to the TR layer from the PW layer.

Now for each of the frequencies weights can be applied identifying to what extent that frequency should be used in the recognition process. Here the weights basically make up the resistance of the receptors where the weights may be influenced by emitting actuators (for instance calling actuator-function) to increase or decrease the weights. The weights should then be a value between *0* and *1*.

By constantly emitting these actuators different kinds of functionality emerge: a delimiter function and a focus function. The weights form some kind of dynamical delimiter for the different frequencies. They don't actually cut off the frequency at a certain volume, but they try to reduce it continually. Just as well by reducing weights to zero (by creating infinite resistance) different frequency

[4](Arons, 1992)
[5](Oghalai, M.D., 2004)

ranges may be suppressed. This focuses on other frequency ranges.

Using the described mechanisms as specific filters for different types of information these mechanisms can be used to tune in to certain sounds, just like with a radio. In order to do so a certain estimate has to be built: an estimate reward mechanism which tries to apply certain patterns to sounds we perceive.

To start with it's possible to shift frequency ranges. In shifting the frequency ranges it's then possible to identify changes in sounds which resemble the changes in speech-uttered sounds. So the changing of frequencies and intensities of frequencies may resemble the types of changes a being would expect when listening to speech.

When these patterns seem to apply, i.e. the likelihood that a certain sound is actually speech is rather high, it's time to find a certain context. But what determines the context?

The context is determined on the one hand by the utterance of specific words and on the other hand by the relation found between these words.[6] These relations will be made up of consequent processes modeling these relations. The processes will have to be freed up, i.e. their resistance of initiation has to be lowered. In case of a larger context this will also mean that more processes have to be freed up. This is harder to do, which leads to difficulties in establishing larger contexts when listening to sounds. This is especially hard in relation with *the Cocktail Party Effect* because of interfering noises. Of course when a certain context is established it may not be the one that was originally sought for. In order to switch contexts the tuning has to continue, just like with finding the right radio station.

The context may be related to different sound sources where each sound source is associated with a specific LO. This LO will then provide phase-shifts. Based on these phase-shifts the LO of a source may be established. Just as well a being may have an estimate of the LO of the sound source. This may then lead to an estimate of the phase shift which may help in identifying the sound source's frequencies and relations between the frequencies for different receivers. In switching sound sources these estimates may also come forth from previously known LOs.

Of course when a voice was previously known the specific memory of the frequency range may also serve as an estimate. By using this estimate to free certain frequencies and suppressing others this may then be used to single out this one sound source.

More generally it's possible to state that usually voices fall within a certain frequency range. A being usually only has to look within expected frequency ranges to find voices.

When a specific sound source has been found it's then possible to focus more and more on this source. This comes forth from on the one hand increasing CO desires and on the other reducing the perception of other sounds by emitting specific actuators. This will also help generating a better estimate which may then lead to identifying larger contexts. In order to do so a being thus needs to become habituated to certain sound sources.

The above principles may be used to solve *the Cocktail Party Effect* to a limited extent. As stated earlier we ourselves haven't evolved perfect mechanisms to do so. We shouldn't try to create these perfect mechanisms. This just isn't

[6](Arons, 1992)

possible.

Conclusion and Recommendations

Reviewing what has been done it now becomes quite clear what is possible using the GToC and how feelings should be fitted into the model. Different ideas, theories and even solutions have come forth in previous chapters which may now be applied to different problems.

In the future many new ideas should be formed based on the GToC which should lead to better solutions. It has quite clearly been shown that it's possible to do so and people should continue to do so.

Bibliography

B. Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society 12*, pages 35–50, July 1992.

C. Barnes. Comedy in dance. In W. Sorrell, editor, *The Dance Has Many Faces*, pages 87–95. Pennington: a cappella books, third edition, 1992.

J. Bongard. *Incremental Approaches to the Combined Evolution of a Robot's Body and Brain*. PhD thesis, Universität Zürich, 2003.

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.

L. Custódio, R. Ventura, and C. Pinto-Ferreira. Artificial emotions and emotion-based control systems. Instituto di Sistemas e Robótica/Instituto Superior Técnico.

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.

C. Gruner. *The Game of Humor: A Comprehensive Theory of Why We Laugh*. Transaction Publishers, New Brunswick, NJ, 1997. ASIN 1560003138.

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. The general theory of consciousness: The abstract definition of the processes required for the emergence of consciousness. Master's thesis, University of Twente, 2004b.

M.P. Mulder and A. Nijholt. *Humour Research: State of the Art*. Number 02-34 in CTIT Technical Report series. CTIT, University of Twente, PO Box 217, 7500 AE Enschede the Netherlands. ISSN 1381-3625.

J.S. Oghalai, M.D. Hearing and hair cells. http://www.bcm.tmc.edu/oto/research/cochlea/Hearing/, April 2004.

R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, Cambridge Massachussets, London, England, 1998. ISBN 0-262-19398-1. A Bradford Book.

The International Synaesthesia Association. The international synaesthesia association. http://www.psychiatry.cam.ac.uk/isa, Februari 2004.

S. Ungerleider. *Mental Training for Peak Performance: top athletes reveal the mind exercises they use to excel*. Rodale Press, 1996. ISBN 0-87596-282-3.