A General Agent Design Specification

Article · January 2004

6 authors, including:


Rieks Op den Akker
University of Twente


Dirk Heylen
University of Twente


Mannes Poel
University of Twente

A General Agent Design Specification

(Master's Thesis Appendix I)

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

Forword

In view of the *General Theory of Consciousness (GToC)*[1] my supervisors wondered how the described mechanisms could be applied to agents. In this report I describe how different instantiations of these mechanisms relate as well as how embodiment influences a being. What I do wish to say here is that it's highly relevant to think about what we're actually doing here. To my opinion only simplified mechanisms should be introduced into agent architectures. So before proceeding with reading the whole of this report I would highly recommend reading the chapter on ethics, namely chapter 5.

[1] (Hobo, 2004b)

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

Definitions

Statements

Chapter 1

Introduction

An important aspect of research is the interaction between agents within a world. In order to interact such agents have multiple systems which need to interact within the agent itself. This may then lead to the agent commencing into interaction with other agents based on its views and knowledge. It's possible of course to derive each of the systems of such an agent separately in order to suit a specific task. But when these systems have been derived, how should they then interact? This thesis is concerned with the interaction of agents and how such an agent may be properly designed.

1.1 The Goals of Research

In order to compose an agent from different systems, first these different systems need to be identified. We as humans have five basic senses which may be used within the world. When looking at robots *other senses* which have been introduced are basically just *simplifications* or *transpositions* of our senses. For instance *radio communication* should be considered as a transposition of the senses because it allows agents to communicate their perceptions. So in order to properly discuss agents it's important to first discuss the senses. These then make up the perceptual system.

When the physical properties limit the agent too much in its behaviour, it's not suited to do its task. Beings in the world as we know it have been evolved to fit their present circumstances. If the circumstances are changed (like for instance with the Dodo) or the beings are put into circumstances they do not fit in, they do not survive. So any agent should have a proper embodiment to suit its environment.

Without a control system the agent doesn't do anything. A being without a working brain isn't anything but an empty shell. The body may be kept alive, but it will not interact with the world. So in order for an agent to learn, decide and interact with its world, it needs a proper control system.

The capabilities we already have now make it highly questionable whether what we can do we should do. It's easy to create something, but why should it exist with no purpose? In order to discuss what falls within ethical boundaries a full set of ethics needs to be derived.

1.2 Approach

Since the main subject is the composition of different systems into one bigger system called an agent, the approach will be focused on the decomposition of systems. Having decomposed the systems, they may then be recomposed. This is consequently done for the senses, the embodiment and the control systems.

In case of ethics instead of decomposing and recomposing a different method of research is employed. Since ethics are highly dependent on circumstances the ethics are composed based on derivations of these circumstances. This also means that the circumstances first have to properly be defined.

1.3 The Structure of the Thesis

Chapter 2 describes the senses as present in an agent. Chapter 3 then describes the influence of embodiment on the agents capabilities. Chapter 4 describes how different control systems should relate. Chapter 5 described the ethics concerned with composing agents. Finally conclusions are drawn and recommendations regarding future work are made.

1.4 On Application

As has been noted in chapter 5 it to my opinion isn't ethically correct to actually implement a being. Instead I propose that only simplifications of beings in terms of simpler agent architectures should be introduced. Although this appendix discusses everything in terms of *beings* here you should actually read *simplifications of beings as caught in agent architectures*. I encourage you to read chapter 5 before any other in this appendix.

Chapter 2

The Senses

There are basically four types of senses. We usually make a distinction between five senses, but in the way an organism should be implemented the difference between taste and smell are less interesting. So we basically have the following subdivision: visual, auditory, haptic and chemical senses. For each of these senses there are some reflexive mechanisms (RMs) and non-reflexive mechanisms (NRMs). These will both be discussed for each of the senses.

Section 2.1 looks at the visual senses. Section 2.2 listens for the auditory senses. Section 2.3 touches haptic senses. Section 2.4 smells and tastes the chemical senses. Section 2.5 determines what happens when a being needs to switch between senses for doing its task. Finally section 2.6 discusses qualia and more explicitly qualitative processes.

2.1 Visual Senses

Within organisms, the most complex and most evolved of senses is the visual sense. In the past many issues regarding vision have been chosen to model. This section proposes a complex vision system containing not just one but many of the features that have been modeled in the past. There isn't just one system called the vision system, there are many different systems working together.

We have RMs and NRMs, both associated with the visual senses. The RMs like the notion of presence of movement are not only present in us, but also in less complex organisms. These systems are even present in for instance flies. We also have associative systems. We can learn to see things that a baby wouldn't. Just as well we attribute certain qualities to our perceived images. These are examples of NRMs.

The networked model of processes this section proposes should generate certain outputs. These outputs in case of actions will here be limited to the vision system. The actions will only be the ones directly related to the vision system.

The RMs which are part of the vision system are directly related to the physical properties of the vision system. What should be our main point of focus should thus be what the properties are of such a system. There are two parts that may be distinguished between:

1. First of all there are the properties of the *Consciousness Reference Model*

(CRM).[1] These properties have been discussed in regard of the GToC and emotions. These same transitions are there in other instances of the CRM. The non-interpretational properties of these systems are the potential of the stimuli and the change of potential of the stimuli. This basically means that there should be RMs associated with the intensity of the picture. Just as well there should also be RMs associated with the amount of change in the picture.

2. Second the RMs are directly associated with motor systems. These motor systems contain physical RMs. These physical RMs may be divided into two systems:

   - **Protective mechanisms.** For instance closing the eyes in case of high potentials.
   - **Focussing mechanisms.** Mostly dependent on tracking and centring on changes in the picture.[2]

For only noting the change of potential it isn't necessary to have a permanent associative picture storage. Different functional systems may be interpreted. These functional systems then only need to consider the change of the picture instead of the true picture itself. This storage isn't actually memory as we perceive it, but it represents a temporal instance of the difference of inputs within a certain time-frame. When the differences are large the RMs do their job and focus on the change of potential.

Sometimes the pictures are stored more permanently in more complex organisms. In this particular case instantiations of the CRM are needed to perform the function of memory.[3] Whether more permanently stored pictures are actually processed changes from time to time, but in complex organisms they can always be recalled from memory afterwards when needed. This however is beyond RMs.

In parallel to the RMs run the NRMs. These NRMs are controlled by cognitive (CO) desire as well as the physical world (PW) influence. The CO desire may express the need to look at a certain part of the PW because this is of some kind of interest to us. The PW potential may be an influence from the PW like the continuous flow of information. An example may be the abuse of this influence by actively applying brainwashing. Having this basic idea of how influences on the NRMs come about one might wonder what NRMs there actually are.

There are certain associative systems determining what to look for. These will also control the motor systems which determine a certain focus. When a focus has been chosen other associative systems determine the boundaries of focus. Within these boundaries it's then possible to explore. When a certain focus has been found a relation has to be established and thus a representation (RE) has to be built of the found relation within the boundaries of focus. Just as well a more global RE may be built describing more global relations using

[1] (Hobo, 2004b, Chapter 6)

[2] A mechanism like this has to some extent been developed by Eggenberger Hotz et al. (2002).

[3] (Hobo, 2004a, Section 4.3)

the memory system. This will then also include relativity depending on locality (LO). More elaborately this leads to the following mechanism (figure 2.1)[4]:

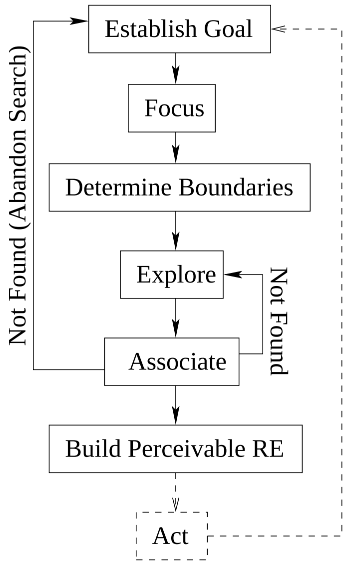

Figure 2.1: The succession graph of the vision-processes

1. **Establish goal.** When NRMs come into play a being has a certain goal it wishes to fulfil. This basically means that it wishes to find a certain relation and it starts by creating the CO focus on this relation. This basically means freeing up CO processes which should form the relation.

2. **Focus.** When the goal has been established, the NRMs may try to overrule the RMs in determining the point of focus. This is done in a recurrent relation with the establishment of a goal, because the focusing process constantly tries to fit the point of focus into the freed up relations.

   What may help focusing on a certain goal is the specific architecture used for the vision system. For the perceived areas of focus it's interesting to note details and the colours of every detail. In areas of less interest less detail will suffice. For instance the human retina contains much more information in the centre than on the outside of the visual range. The reflexive focussing on objects has been vastly studied.[5]

3. **Determine boundaries.** When a point of focus has been selected the surroundings may be scanned in order to determine the limits of the relation which has been focused upon. When for instance looking at a specific other being specific parts of that being may be focused upon (like the face, if it has one). In order to do so certain edge detection mechanisms should be in place.

[4]A process only initiates the succession of one associated process in the direction of the succession graph.

[5](Sandini et al., 2000, ed.: Zanker and Zeil)

What I would now suggest is not installing edge detection mechanisms in terms of drawn edges. The boundaries aren't interesting, the surfaces are. What I would now like to suggest is that instead of the classical edge-detection mechanisms there should be surface selectors. When a surface internal to a certain set of edges has been found, everything outside of this surface should be repressed.

In order to develop more detailed divisions into surfaces the being should now develop multiple associative processes each new process providing for extra detail. Every process selects one surface. When looking for instance at an infant, the infant at first laughs at anything oval. Then at anything with a smile. The details are added bit by bit, until the infant has a clear picture of its parents in its head. When it has this clear picture, it will not laugh at anything else. This is illustrated in figure 2.2. The way this works is directly derived from the GToC.

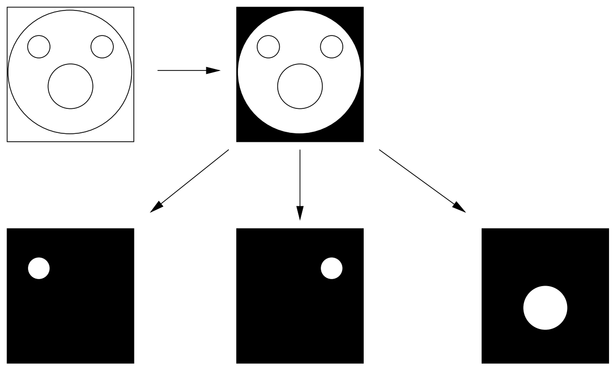

Figure 2.2: Picking surfaces in the face

So in order to simplify mechanisms I suggest that instead of determining boundaries and searching within these boundaries, the surface within which should be sought itself should be provided. After the surface has been provided, this surface may be searched. An object is now represented by a collection of successive divisions into surfaces.

4. **Explore.** When the surface of the chosen focus has been determined it is then possible to explore the relations focused upon within this surface. This exploration will then be limited to the relations which are made up by the freed up CO processes into which the relations should fall. One way the exploration might now be imagined is that a group of smaller surfaces regarding their form and constitution now fall into these relations.

5. **Associate.** The association comes forth by actually determining the sought for associations within the current view. Of course it's also possible to establish that the relations do not hold within the current view, which may mean that a being will have to look further or maybe after some time abandons the search.

6. **Build perceivable RE.** When the relations have been established and the point of focus has thus been fitted into the previously freed up relational structure, a perceivable RE of the found relation may be built. This

perceivable RE is generated by the association within the CO structure which induces the emitting of the perceivable RE and may be associated with the principle of qualia. This perceivable RE may then be perceived itself and associated with the perceived inputs.

When these associations within a picture are to some degree experienced again a new association may be drawn. For instance, after having been told many times what a car looks like, a child learns to distinguish cars more generally instead of just one specific car being a car. This means that an association exists between associations found in single images. Or, to put it differently, by fitting associations found in a point of focus into fuzzy boundaries, the associations found may be classified. So now we don't just have direct picture memory, but also associative picture memory.[6]

The following example may serve to illustrate in what way we should look at the operating of the visual senses. In recognising things, the image size changes with the distance. So when trying to associate with something it's probably best to speak of relative sizes. Whether something is big or small will then just be another association. This may then lead to for instance the distinction between a real car, a toy car or maybe a car seen from a distance. Here the distance can be established by for instance the amount of change in relations when only minimally shifting the foci. Then a toy car will have a large amount of change in relations because of perspective and a far away car will only experience minor or non-noticeable change. This to a large extent influences the RE of what we see.

2.2 Auditory Senses

The sense of hearing in its basic make-up isn't a complex type of sense, but still it can be used to communicate very complex messages. Here the complexity arises from a large number of smaller sounds which may together form sequences with a certain meaning. To protect ourselves and our ears there are RMs and to communicate there are also NRMs.

It may happen that certain receptors in our auditory system may receive too many stimuli. In case of our eyes we would just close our eye-lids and reduce the diameter of our pupil. Although parts of our ears just like any other kind of receptor may grow a bit numb for certain frequencies, this doesn't mean that it protects the senses themselves. In case of our ears there are little hairs which shiver when certain frequencies hit our ears. When sounds grow to loud the hair-cells of these hairs die.

Suppose the same hearing system as present in humans were used with a robot or maybe a microphone was introduced. If this were done it would mean the same thing when a sound became too loud: the sound would destroy the sensory organs. To protect the sensory organs RMs should be in place. This would mean that an implementation would have hands that can cover the ears or just ear-flaps which could close. You don't want to install permanent ear-plugs, because for general protection it's still needed to perceive a bang as loud as a bang.

[6](Hobo, 2004a, Section 4.3)

Although architectures of less complexity may have less complex communication systems, it's interesting to take a peek at things like natural language. Natural language comes forth from a sequence of sounds. Each sound on its own doesn't necessarily have to imply an association. The different sounds together however will indicate relations and can be deduced step by step to form that relation.

So what you would now get by analysing a sentence could be that there are specific parts of networks associated with establishing best guesses of different grammatical conventions. This would then lead to a likelihood function of what might be the subject of the sentence, what the subject does and many more associations. All these best likelihoods may lead to an interpretation of the sentence. As a subsystem a similar likelihood function has to be trained to make best guesses regarding the meaning of single words in a sentence. Then an association may be drawn between those words which defines the sentence's meaning.

A more primitive version of this type of communication is music. Here music is mainly concerned with establishing relations between sounds without attributing meaning to these sounds. Funnily in music terminology like *phraseology* is used to indicate the relation between different parts played. This way, musicians may for instance enter a *dialog* with each other. One musician may also play a dialog representing the question and the answer entity on his own. This then leads to the *solo* or basically a story which he tries to tell the audience. In describing natural language it may be better to first learn a certain mechanism to describe these differences based on intonation using music. Then this may be applied to natural language in correlation with the meaning of words.

To implement a sophisticated hearing mechanism multiple small mechanisms will now have to be implemented. These are related to the following two general mechanisms:

- **Intonation mechanism.** This will then serve to make best guesses of the dynamics of a sentence.

- **Association mechanism.** This will have to associate with words, sentences and meaning. The words should be contained in some kind of extendible vocabulary.

Although these two basic mechanisms suffice in case of statical sentence-recognition it's necessary to create an even more specific hierarchy. This hierarchy would then be the hierarchy as present in us. This hierarchy will first have to establish the meaning of loose words which may then form sentences. This way it's possible to form more dynamical sentences.

In case of word-recognition mechanisms a subdivision can be made into three mechanisms (figure 2.3):

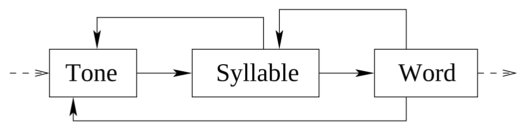

Figure 2.3: The succession graph of the recognition of words

- **Tone mechanism.** This should be concerned with how sounds come into being by looking at intonation. This will basically form the alphabet of a language.
- **Syllable mechanism.** Based on pauses and abrupt differentials in change of tone the different syllables may be found.
- **Word RE.** When the different syllables have been recognised best guesses will have to be made in putting them together. This will then lead to loose words with separate possible meanings.

Of course sentences may be derived in the same manner from words. In case of sentence-recognition mechanisms the following three mechanisms may be recognised (figure 2.4):

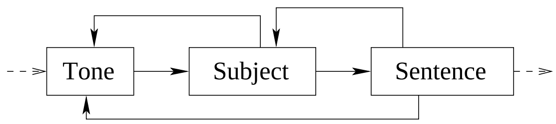

Figure 2.4: The succession graph of the recognition of sentences

- **Intonation mechanism.** To each word should not only be attributed a meaning, but also an intonation which may help determine the subjects and meaning of the sentence.
- **Subjective mechanism.** This should indicate the role of each word within a sentence.
- **Sentence RE.** The sentence RE should basically provide an image of what is meant within the sentence. This is needed to for instance establish that although it's stated differently two phrases may have the same meaning.

The word and sentence-recognition mechanisms both contain a RE builder. It's not possible to describe all classes of sentences. What it should basically do is form a general RE function which may also consider the equation of representational content of the sentence. This representational content should be determined by the general relations which form a memory of the sentence. Multiple sentences should then possibly be fitted into this general relation, where the relation determines the equality of meaning. The equality of meaning then determines whether two sentences both occupy the same general RE.

Whether two sentences both occupy the same general RE is largely dependent on the listener and thus is just an estimate. This again stipulates that every relation is determined based on the highest likelihood. The process of determining the sense of a sentence's meaning is a very explicit example of the mechanism.[7] In order to build more dynamical speech recognition software this should be taken into account.

Of course beings are also able to recognise other sounds other than sentences. Because these sounds are just loose sounds they may be regarded as *words*. From

[7](Hobo, 2004a, Sections 3.4 and 5.1)

the possibility of finding associations with words and sentences then automatically follows the possibility of finding associations with other sounds as well. Because sentences are the most complex cohesion of sound we know and a word can be seen as the most simple type of sentence, loose sounds are recognisable as well. The only distinction then is that a *word* is a RE in sound of an identifier which we have attributed to something.

Loose sounds aren't REs, but they are effects which may be identified. These identifications could in some part be related to synonyms. Basically an identifier is a synonym to the thing it identifies, just as it is a synonym to any other identifier with the same meaning. But the way we use the word *synonym* only relates different identifications of meaning within natural language to each other.

2.3 Haptic Senses

Haptic senses form quite a complex system: heat and cold; tickling and itching; pressure and pull. Based on the different types of properties we may attribute certain values to an estimated internal state of an object. Just as well in combination with our own internal change during the sensing along the contour of an object this may lead to an estimate of exterior properties of an object. The sensing of internal change also means having a sense of direction based on for instance gravity-pull towards a large object. But these are all very advanced mechanisms. What basic mechanisms lie underneath?

To start with there are again RMs. When something produces too many stimuli we tend to draw away from it. For instance when you get stuck your first reaction usually is to pull. A good example of a game which utilises this is *the Chinese finger trap*. What basically happens is the following. The trap is a simple tube which you stick your index fingers into. When you try to pull your fingers out the tube tightens and you are not able to. In order to overcome this problem you'll need to apply logic: if you pull it tightens, maybe it loosens when you push? This is indeed what happens. By pushing the tube opens up and you're able to pull your fingers out. This game uses the natural reflexes which you'll have to overcome with common sense. Just as well when too much heat or cold is applied this will also induce pain and you'll try to pull away. In case of too large an itch it will feel like something is eating your hand away and you'll not only pull it away but also try to scratch whatever itches off. In some animals the result of receiving too many of some of these stimuli is actually that they just bite of the limb to save themselves.

Another RM which may be noticed is shivering because of tickling feelings. This may be associated with for instance making sure flies don't sit on you, just like a horse or a cow that shivers and also wags his tail to remove the flies.

Touch may also be used in a more complex manner to determine shape, cohesion and the type of surface. This will mean that NRMs will have to be implemented to associate with different objects and other relations like mechanics.

The most complex aspects of objects are not determined by heat or cold or by tickling and itching of the surfaces. The most complex aspects are basically the amount of resistance objects provide in pushing against or pulling away from the object. These two principles can be subdivided or are related to five other principles which may be perceived:

- **Cohesiveness.** For instance water vs. sand. Water wants to stick together, sand doesn't.
- **Massiveness.** For instance water vs. sand. Water is made up of one type of particle where all the particles cling together. Sand is made up of many perceivable particles separated by some gas or fluid.
- **Protrudability.** For instance carton vs. rock. Carton can often be protruded with a finger, rock can't be protruded with a finger.
- **Solidity.** For instance rock vs. water. Rock can be put down anywhere without loosing its shape, water splashes.
- **Adhesiveness.** For instance water vs. paper. Water sticks to your fingers, paper won't.

Based on the five mentioned principles it is now possible to form a relation. This determines the view of the world concerned with what place objects take up in the world and how they relate to each other. For some part this is also correlated with movement to establish the change of relations over certain distances, but this is beyond the present context.

This section as previously stated only serves to discuss the make-up of senses. It should also be noted that the general posture of a being can be noted by the being itself by determining the above properties within the physical make-up of itself. I'm not going to delve any deeper into the more far-going consequences of what these senses can do.

2.4 Chemical Senses

Taste and smell are both very definitely correlated, partly because smell influences taste and more importantly because they both use similar mechanisms. Associated with taste and smell are RMs to prevent any harmful behaviour. Just as well it's possible to identify specific tastes and smells, which may be singular or composite, using NRMs. Both the RMs and NRMs will be discussed in the here mentioned order.

Taste and smell are both senses specifically meant to distinguish between the good and the bad of food, drink and air. They are brought about by the dissolving of particles into a layer of fluid after which the dissolved particles may attach themselves to receptors.

The RMs influence mainly whether we explore further or turn away. When the receptors receive too many stimuli in case of taste we spit it out and in case of smell we turn away and maybe stop breathing for a while. Just as well something that tastes or smells good encourages to explore further.

The establishment of a certain taste or smell comes forth from NRMs. In case of any kind of consumption it's highly important to taste or smell it properly. The taste and smell may then be associated with specific objects, surroundings and experiences. For instance when walking through a flowery field and consuming the air, the smell of flowers and other greenery may be identified.

Regarding taste and smell, the tastes and smells are composed of multiple instances of basic tastes and smells. The way they are perceived for each basic taste and smell a being has specific chemoreceptors which will trigger only specific reactions when stimulated. The actually perceived taste or smell then comes forth from many of these basic tastes and smells being stimulated in larger or smaller amounts. This way a composite taste or smell is derived.

Basically chemical senses are the most straightforward kind of senses. Beings only deal with single instances of tastes and smells at the same time. There is no preconceived grammar which creates a certain kind of language. Of course it's possible to influence following perceptions, because certain receptors will be freed and blocked. This will make sure that some tastes and smells are perceived stronger than others. So in for instance composing a meal it's possible to build a preconceived succession of tastes and smells. Here every previous taste makes sure the next is even better.

2.5 Exchangeability of Senses

All the senses which have been discussed until now form rather complex systems. It is questionable whether beings actually need all these senses in order to be able to manoeuver themselves through their worlds. Quite some functionality partially overlaps for each of the senses. I wish to argue here that it's largely world-dependent whether beings are able to survive or not with limited senses.

Within our own world there are quite some cases of human beings who have lost one or more of their senses. Most commonly noticeable losses are the ones occurring in case of deafness or blindness. People who loose these senses are most dependent on vast social networks which form within society in order to get by properly.

In case of blindness an obvious physical reaction occurs by the natural enhancement of other senses. This is quite a logical adaptation of the brain, because parts of the brain which are normally associated with vision are now free to perform different functionality.

Many physical aids have been added within society to make sure blind people get by, which indicates a direct form of social adaptation. In order to get by in traffic sounds have been added to for instance traffic lights and special sticks have been manufactured to compensate for the loss of sight. Also different markings which can be sensed by blind people have been added to for instance train stations to indicate routes which can be followed by blind people. An instance of such a marking are the ribbed tiles which indicate the walking direction on platforms at train stations in the Netherlands (and probably also other countries).

In case of deafness a necessary change within communication principles arises because it isn't possible to focus on regular speech patterns anymore. The loss of sound needs to be compensated for by introducing more visually oriented symbols as well as lip-reading.

Where in nature animals would sometimes die because of the lack of certain senses, social structures which have grown more and more complex form a solid back-bone to make up for these losses. Physically a being may adapt to a certain loss of senses, but a being isn't necessarily thus adaptable that it may survive. The exchangeability of senses largely depends on the super-imposed task of a being within a certain social hierarchy or structure.

The only thing that should still be noted is that for instance a deaf animal in certain socially isolated areas may quite easily survive. In some cases blind animals don't have too many problems as well, as can be seen with for instance the mole. Lot's of the social structures which have grown within our society come forth from the direct need created by the structure of society. In a forest without cars a blind person will have a hard time trying to be hit by a car. This basically means that there will be hardly any traffic lights with beepers. On the other hand he's much more susceptible to being eating by a bear, because it's hard to listen properly and not to run into one of the many trees. It is however usually so that the hearing of such a person is much more sophisticated and the person will thus still have possibilities to survive by preventive evasion of danger.

2.6 Perception of Senses

When using any kind of sensor within computer science, the incoming data is parsed to a value suitable for computation. This value doesn't mean anything to the computer in terms of good or bad or associated feelings. The value is only used for basic computational operations within the computer. What we as people claim to do is to actually perceive the senses. So this means that an image of the senses needs to be built internal to our being. This image may then be perceived again by us. This image is made up of what we like to call *qualia*.[8]

Although we do have some basic habits, there are also some habits which we learn during our life-time. These habits are associated with what we like to call taste. This taste is not just associated with food or beverage, but also with art, science, quality of life and many more things we tend to enjoy. Where do these new habits come from?

An example might be getting used to some tastes and smells in relation to food. This happens because we know that the later positive reward will be larger than the first few negative rewards added together. Consider eating olives or drinking certain wines. In order to enjoy them, you need to learn to enjoy them. This way we experience qualia associated with them which may be identified as *good* or *tasty*, to put it in terms of eating. This means that we learn to associate positive feelings with stimuli with which we formerly associated negative feelings. So developing any kind of taste is associated with some kind of NRM which ascribes a certain result to different tastes. This may for a large part be inherent to culture and social status or the need to fulfil such a position. This of course doesn't mean that certain tastes are more sophisticated than others, we just claim them to be so. This can be indicated by that even taste is subject to fashion and after a while becomes common instead of sophisticated. A large part of our taste is determined by what we remember to be sophisticated.

Positively rewarding certain states because of the knowledge that future rewards will be higher is an important part of reinforcement learning. This may happen even though intermediate states normally are experienced as being negative at first. Funnily this indicates that qualia are mathematically induced. The feeling we have that arises from tasting something is largely influenced by external factors. So in order to become habituated to something there should be some basic reinforcement learning mechanism. This is also logically plausible

[8](Hobo, 2004a, Section 2.3)

because food, drink and air are directly linked to our survival. Considering the influence of what we see and remember these qualitative systems most probably reside in or are directly related to memory.

Chapter 3

Embodiment

Within artificial intelligence, psychology and many other fields of interest dealing with consciousness, many types of physical architectures have passed our visionary screens. What if a being *looked* like this? What if a being would *be* like that? What a being *looks like* and what a being actually *is* although not always giving away information about its character does usually ascribe him a certain task. For instance in a scary movie the girl with the big eyes that constantly clings to some other person on average has the very clear task specification of screaming somewhere during the movie.[1] But how do architectures more generally influence tasks or how do tasks more generally influence architectures?[2]

Multiple parts may be discerned regarding the architecture of beings or agents. In order to discuss these parts section 3.1 first discusses the constraints that may be posed on the architecture. These will comprehend the environmental constraints from which also come forth the task-related constraints. An important aspect of exploration is propagation. Section 3.2 discusses different types of moving through environments. In order to properly explore, some kind of representation of the environment is needed. Section 3.3 discusses perception. Finally beings or agents are or may be possibly be made up of different materials. What influence such materials may have is discussed in section 3.4.

3.1 Constraints

What a being actually can do is very much related to the environmental constraints. In some fantasy worlds quite a lot of rules may be bent or broken, but in reality the rules are very strict and any kind of being just has to learn to live with them. These rules are basically the physical laws or the mechanics and dynamics of a world. Based on the location and the corresponding surroundings of a being, a being will have to adapt in order to survive. So, what environmental constraints may be identified?

To start with, a being will be immersed in some kind of material. This may be gas, fluid or some kind of mixture of gas, fluid and solid material. For instance a worm moves through dirt, which is a mixture of sand, water, air and

[1] I don't think this is an example I came up with.

[2] For a full treatise on embodiment I suggest you read Pfeifer and Scheier (1999).

nutritious supplements. It filters the nutritious supplements out of the earth and excrements whatever it can't use. A fish swims through water, from which it filters air. Within the water it feeds on other solid nutritious supplements, which may be other beings like plankton.

Another constraint which may be imposed on a being is the amount of space it receives to manoeuvre. Take for instance some types of white blood cells. These have adapted to not only move through the veins, but also between the cracks in the walls of the veins. Just as well a small snake can more easily crawl through a narrow crack with many corners than for instance a mouse or a rat.

Just as well other objects which are placed within the world have a certain form and shape or dimensions. A tree for instance grows into the sky and some beings may choose to try to climb into these trees. Other beings may decide to use the trees for other purposes, for instance to chop holes in them for nesting areas.

So there are three basic constraints concerned with an environment. The environment has physical laws, which hold at all times and can't be broken. A being can be said to be contained in some kind of conductor through which it has to navigate. A world is basically made up of objects with certain dimensions and material composition which may sometimes be used as conductors when the being is fitted to do so.

The task of a being basically is what it has to do to survive. It has to maintain its own architecture, making sure it doesn't become a lifeless object itself. In order to do so it will need some kind of energy source. This energy source then makes sure that it can keep up the processes which actually maintain the energy source. This brings the energy source in a recurrent relation with the physical task a being has to uphold.

Of course the structure of a being deteriorates after some time. A being doesn't live forever. This means that it will either need replacement parts in order to ensure the continuation of processes. This will in time lead to a being fully renewing itself. Another thing it may do is procreate creating offspring which in their turn may ensure the continuation of processes. This may then also lead to changes in the architecture through mutation of some kind, which in turn has been named evolution. This way the world is renewed over and over again.

So the task of a being is to keep life alive within the world. In order to do so it has to adapt to the environment or find a piece of the environment which it is adapted to.

3.2 Propagation

Different beings propagate through the world in different manners. Some beings fly, some beings swim, others crawl or walk or dig or do other things. Basically a subdivision may be made between the ways beings propagate through the world in relation to the material they are using for propagation.

A being may propagate through a world where it has to uphold a certain amount of *lift* in order not to sink or fall. It is then surrounded by some kind of liquid or gas which it has to manoeuvre through. This is what we call *swimming* or *flying*. A fish on the one hand has to create certain amounts of upward lift and on the other has to push itself forward. A bird on the other hand can often only remain in the air by gliding forward. There are of course also birds which hover in the air, but quite often this isn't possible. The way the forward motion is now created may be by using muscles, but it's also possible to create a jet propulsion system to propagate through the air. In order to change direction as well as go forward any kind of combination ranging from only muscle strength to only jet propulsion systems or a combination of both may be used. This highly depends on the task of whatever is created.

A being may also have a *surface* onto which it is pushed by *gravity*. Here the being has to lift itself up from the surface to a certain amount to be able to engage into a motion on this surface. This we call *walking*, *crawling* or *driving*. Walking or crawling may be instantiated by using different amounts of legs. Driving may be used by using wheels like for instance with cars. Combinations of the two may be thought of, where the legs are then for instance used to push instead of walk and the wheel just rolls.[3]

A being which *clings* to a surface which *gravity* tries to pull it loose from is said to *climb*. Here the being needs to make sure not to fall. It will basically need to grapple on to the surface. This may be done by using hook-shaped or grappler artifacts which may or may not be part of the beings body. Just as well in case of smooth surfaces it may use suction to stick to the surface. A snail uses suction. A fly uses microscopic hooks. When a human climbs in a tree it basically uses its hands as grapplers to hold on to the three.

A being that moves through some *solid* material creating holes is said to *dig*. A mole or a worm uses digging for propagation.

Which has become evident is that not only the type of relation, but also the type of material a being is in relation to highly influences its possibilities regarding motion. For instance when on plain land humans usually don't have any problems propagating themselves. But when a human is on ice the chances that the human will fall are a lot higher. Another thing that may be noted is that one type of motion doesn't prohibit another type of motion. So the type of motion should be suited to the task. A being may now choose a type of motion in order to reach a certain goal. For instance human infants primarily learn to crawl to propagate. In order to propagate faster human infants learn to walk and when having learned to walk they learn to run. Maybe a more obvious type of motion not common to us is swimming. This we also need to learn.

As now has become clear within nature there are quite a lot of motor systems. We ourselves even invented some new ones like the ones controlling the wheel. The invention of the wheel itself is a very simple invention which permits us to easily move from one spot to the other. Our limbs however, as we have come to discover, can be much more complex in control. Why is it then so easy for us to move around? From current studies follows that we should actually consider systems like the ones controlling the wheel to be the most complex of the different motor systems. Instead we should use (and not necessarily constantly control) motion by getting the most out of physical laws. This also means that the robots morphology largely depends on its motion control system and vice versa.[4]

Let's first look at this one discovery of ours: controlling the wheel. The movement of the wheel has quite some parameters. It's associated with a certain

[3](Tokyo Lectures, 2003)
[4](Paul and Bongard, 2001)

number of rotations per minute, it may speed up and it may slow down. This then corresponds to an established speed within a world. All these parameters have to be constantly monitored in order to control the speed. When this isn't done properly the speed dances up and down and there is no possible way to properly make an estimate of the needed action in future situations. This usually slows down the system a lot. But why then is it easier to control a properly built walking system?

In case of a wheel the motor is largely part of and controlled by the propagated subject. In walking the subject makes use of the motor system as contained in the physics of a PW. This means that the constancy of speed is established by the physical properties of the subject. More particularly the constancy of speed is attained by the make-up of its legs and the propagation of its centre of mass.

Where the wheel has to be controlled all the time, the legs only have to be controlled within specific time-frames. What's just needed is the following three controls:

1. The pushing of the leg,

2. the letting the leg swing forward after it has been lifted and

3. the putting down of the leg.

Of course in order to determine the direction certain angles of the legs during each of these actions may be changed. Just as well the speed may be controlled by the applied force during specific time-frames.

In order to have a properly walking system it needs to remain in balance. Taking humans as an example, in order to remain on our feet other parts of our bodies start swinging to compensate for the change of centre of mass induced by walking itself. Normally our hands and arms start swinging. When we put our hands in our pockets our shoulders start swinging. But is there any motion control associated with this? By just letting the arms swing loosely, physics *for a large part* makes sure that they swing in the right fashion to establish an approximate constancy regarding the centre of mass. So there's only *limited control* associated with the arms during the walking. They do tend to grow tired, but personally I can't say whether this comes forth from control by the muscles or perhaps resistance by the muscles. That they also start moving by themselves without control follows from the laws of inertia as proposed in classical mechanics by Newton. A simple thought experiment of a robotic structure with two legs and instead of arms two ropes attached to the shoulders with weights attached to the end of the ropes shows how this will work. Of course this fails when the robot not gradually but suddenly stops walking, because then the arms continue swinging. This leaves the robot off-balance. This may also happen when suddenly starting to walk instead of gradually.

The control of most of the groups of specific motor systems is now limited to specific time-frames or almost instances in time. This greatly reduces the workload of CO processes. In case of most of the possible movements this is actually precisely what we're after.

Since in walking a subject needs to watch out for many other things as well it can't be constantly occupied with the movements within walking. In the past the way the motor systems worked within robotics they didn't use the dynamics of the robot but tried to control everything. This led to utterly slow robots. Nowadays research also focuses on using physical dynamics of robotic structures to make it move. But should this be the only system that we're after? The answer quite obviously is *no*.

In some cases the subject does wish to constantly apply force or maintain a relatively constant movement. But these aren't everyday movements but specific precise movements which the subject has to focus on to carry them out correctly. For instance in navigating through a narrow crevice where we don't want to stir rocks which may fall upon our heads it's important to move carefully. But here the main focus of attention is the movement and small limited parts of the world. In everyday walking we are more concerned with determining a general route, seeing where we are. This asks for a more general observance of the world around us. Or maybe we are thinking within ourselves, which may of course sometimes lead to bumping into things because our movements aren't well-controlled.

In creating a motor system it should be well established that there are certain limits to the required precision of movement. When movements require a less perfected control and can depend more on the dynamics of the subject, it should be accepted that the control system may not work properly in some cases. Here the higher level of control should come into play where the main focus is the controlling itself. Nothing else would in such a case have to interfere. In case of using the dynamics the area of attention is much broader. The subject doesn't need to focus as much on its precise movements, because its general movement serves him well enough. The constancy of for instance speed then largely follows from the dynamics of the system.

To establish a correctly working motor system it should have certain physical properties to be able to do the above.

On the one hand the motor system should be able to fully relax. The physical properties of the system and the physics of the world may now do their work in the movement of the subject.

On the other hand it needs to be able to apply force. This is needed to grab something or speed up and even a little to maintain a certain constancy of speed. Of course there are many other things as well where the application of force is necessary. This may also be used to interrupt a certain movement, which may be dynamically or statically controlled.

The motor control system should go through a certain developmental process. This process has been explored by for instance Lungarella and Berthouze.[5] Their research focuses on the freeing and freezing of degrees of freedom in learning certain motor coordinated skills.

Taking a few steps back to the wheel, it's of course also possible to make this system work. By introducing a cruise-control system in the beings make-up it will only have to utter control commands at the times that specific controls are necessary to manoeuvre properly. The rest of the time it may then depend on the correct operability of the motor-system itself.

The above conclusions can be applied to any type of motor system. The motor system needs to contain these properties to a certain extent relative to the environment it's in. Knowing this using different materials, modalities and morphologies different types of beings may be created.

[5](Lungarella and Berthouze, 2002)

3.3 Perception

Perception has been part of our existence since we were born and seems a natural and obvious aspect of our being. We feel that we didn't have to learn to perceive and it just came as natural to us. This can partly be said not to be true, because we still need to learn to deal with the signals we receive. But this doesn't change that all the architectures are already there and we can just use them. In designing any type of agent sensors are probably the first issue which make an agent tall or fall. There are three aspects of sensors and the data they provide which basically make or break a robot.[6]

First of all there's *sufficiency of data*. In order to properly deal with the environment the sensors should provide sufficient data. This will to a large extent influence the behaviour of the agent. Suppose we build a robot with only touch sensors. This may then only operate at low speeds. When it bumps into something it may often have to make a turn. When a robot is operating at high speeds it will only behave like a crash test dummy once and then seizes to exist. In order to operate at high speeds more sophisticated mechanisms are needed. Most probably this will be some kind of vision system, because this contains a lot less noise than for instance auditory systems.

Second there's *computability*. Although sufficient data is needed, the provided data shouldn't be to complex. This will ask for too high abstraction levels, as well as take way too much computation time. For instance when considering a camera system, often high resolution camera's are provided. These contain way too much data. The resolution needs to be manually reduced by for instance taking averages of square areas in the picture.[7] This way the resolution is dramatically reduced.[8]

Third there's *physical efficiency of the sensors*. The physical efficiency of sensors is basically defined by size, power-usage and modality. The size greatly influences the size of the robot and thus its manoeuvrability. Large cameras may contain for instance more sophisticated zoom-systems, but are harder to handle. The power-usage of a sensor often shouldn't be too high, because the agent quite often should be self-sufficient. In order to be so, the agent should be able to operate at low power-usage, needing only small batteries. This makes sure that it can last longer and it also makes the agent more economical. The modality of a sensor means that a sensor should be easily replaceable in case the sensor doesn't work properly or the designer wishes to experiment with other sensors.

3.4 Materials

Having considered what the basic necessities are of a robot, it's then possible to choose the proper materials. These material will have to follow the guidelines which have been mentioned above.

The material should adhere to the environment's constraints. Take for instance a robot which has legs made of metal with low melting temperatures.[9]

[6] These are generalised principles of the ten principles of Pfeifer and Scheier (1999).

[7] (Pfeifer and Scheier, 1999)

[8] According to (Pfeifer and Scheier, 1999) it may also be a good idea to do this with lower quality camera's in order to reduce noise by averaging it out.

[9] (Tokyo Lectures, 2003)

The way it moves is by deforming its legs to suit its propagation task. This can only be used in environments where the maximum temperature is below that temperature. Otherwise it will not be able to operate properly because its legs are soft all the time. A solution may partially be found in sufficient cooling systems, but this will only be appropriate for certain temperatures. Of course metals with a higher melting temperature might be used in hotter conditions, because the base-temperature is higher and the supplied power will be lower to achieve the same effect. This is a very clear example of how world conditions may influence operability.

The material should adhere to the imposed task by the environment. Sometimes a robot will have to manoeuvre a lot through crevices. In case of hard materials this may pose extra problems because they easily get stuck. This can for instance be solved by using small wheels on the outer surface of the body which basically provide a smooth conduction of the bodies movement. In case of not too harsh environments for instance gel robots[10] might be used which have very specific properties. In sea life nature has provided us with creatures like jellyfish which illustrate an instantiation of this kind of beings.

The morphology of a robot greatly influences its operability. The placement of its limbs and sensors obviously is very important. For instance with only touch sensors on the buttocks a robot would constantly have to sit down in order to explore by touch. This is a rather extreme example. Less extreme would be stating that with touch sensors only on the inside of the hand and not on the end of the tip of the finger it's not possible to feel what's at the end of a very small hole. So basically it's very important to experiment with different morphologies for different tasks. Quite often these can be developed using artificial evolution and simulation.[11]

In current research especially but also in previous research modality is also a very important aspect. In the past it was mostly important to be able to experiment with different morphologies yourself and in order to easily replace parts. It should be easy not only to upgrade but also to repair. A simple example would be designing a new battery which lasts longer, but still fits the previous fitting of the old battery. This way the battery can be easily replaced. A nice example of current research would be self reconfiguring robots which reconfigure to suit their tasks.[12] Different modules basically learn to work together which may evolve into highly complex structures, where every module now gets a very specific function.

What may be described as a general trend in embodiment momentarily is automated reconfiguration to suit specific tasks. In the past the embodied architecture was mostly superimposed by its designers. Now architectures may be formed by letting reconfigurable units learn to transform into a certain structure. This is usually learned in simulations to speed up the process. Just as well genetic programming may be used to create new agent architectures. To a large extent this is how it should be, because we aren't as efficient in designing architectures as processes like those found in nature. One of the most important aspects of natures ingenuity is the fact that solutions are found by chance. More appropriately stated, they are found by letting the most suitable architecture continue its processes.

[10](Otake, 2004)
[11](Bongard, 2003)
[12](Kamimura et al., 2001)

Chapter 4

A Control Architecture for an Agent

In previous chapters lot's has been said about the GToC and derivative ideas. What would now be interesting is seeing how the relations are formed between multiple instances of the CRM and the senses to work together as a being. This is done by designing a control architecture (not considering the physical make-up) for an agent.

The chapter has been divided into a few parts. Section 4.1 discusses the general architectural constraints of an agent that would have to act in a world like ours. Section 4.2 describes how tasks may be identified for a specific agent. Section 4.3 describes which systems need to be implemented to control the agent. Section 4.4 concludes this chapter with a few remarks regarding the implementation of an agent.

4.1 General Architectural Constraints

In considering the architecture of an agent there are quite a lot of constraints that will have to be dealt with. Each of these constraints will to a large extent define how the agent will behave or, more specifically, what his behavioural constraints are.

In order to have a functional agent it needs an environment to act in. The environment will not be discrete, i.e. it doesn't behave according to a fixed number of states. In order to be able to interact with such an environment the agent will thus need to be flexible regarding its actions. (In many cases this may mean such a simple thing as not always fully turning right or left, but only partially.) The actions and environmental states may to some extent be classifiable, but the state-spaces need to be extendible in order to be fully able to cope with any given circumstance. Of course with the size the state-spaces may now become, this means that these will have to be virtual state-spaces. The agent needs to be able to generalise its position in the environment and the environmental state and thus reason about its actions.

Reasoning comes forth from making decisions based on estimated consequent states. So an agent needs to be able to deduce which states will follow based on current and previous states. To do so it needs a certain amount of memory.

If it doesn't the agent never will be able to understand that one state follows from the next, let alone influence future states.

In *appendix II*[1] the continuation of processes was indicated to be a very important aspect of being. The continuation of processes may now be related to the run of an agent.[2] A subset of such a run may end with a change in the environment or an action. In case of the control architecture of an agent seen from the perspective of the GToC this has to be redefined.

definition 4.1 (end of subset of control run) *A subset of the control run always ends with the emitting of actuators.*[3]

Whether these actuators are able to do their job then highly depends on the environment which introduces a certain amount of uncertainty into the agent's control architecture. Of course the environment's property of being indiscrete already carried a high amount of uncertainty with it, but this makes it even harder.

The termination of a run of a being is directly related to the termination of being itself. When the being doesn't exist anymore, the control process halts and vice-versa. They are mutually fully dependent.

Agents quite often are modeled being deterministic and behave according to a vast set of rules. Two things can be stated to show that in case of the GToC it isn't possible to do so. First of all an agent is part of the environment it lives in. Second the environment is partly a carrier of the control architecture of an agent. The second issue indicated that the internal structure of the agent is non-deterministic and thus the control architecture is also non-deterministic. So the control architecture should be a highly flexible one.

As can be seen the internal structure of the agent now determines the reactiveness of the agent. An agent which doesn't take history into account according to Wooldridge is said to be *purely reactive*. The single principle of a memory being present shows that it's not a purely reactive agent. One could consider a simplification of the agent by removing an active memory system. The whole internal structure even without including the memory of the agent designed here is dependent on its history because of learning. This means that the agent itself never is a fully reactive system. Learning then doesn't mean that it has an *active memory*, but the agent can be said to have a *passive memory* learning from previous experience and thus states.

Now Wooldridge clearly states that an agent has very simply identifiable behavioural mechanisms. It has a function to *see* and a function to generate an *action*. The first builds a perception from the environment. The second decides on an action based on zero or more perceptions. The environment thus has two interaction points, one for the perception of the agent and one for the action performed by the agent. In the CRM these would be respectively the receptors and emitters. The difference here is that the physical build of the agent is part of the environment in view of the GToC. So the action is generated outside of the agent control architecture, stimulated by actuators. The perception which is fed to the CO process to create CO desires is built within the processing functionality itself.

[1](Hobo, 2004a)
[2](Wooldridge, 2002)
[3](Hobo, 2004b, definition 5.25)

The basic architecture as proposed by Wooldridge is a bit distorted in terms of the GToC. Actions according to his model are performed *on* the environment. According to the GToC they need to be seen as being performed *within* the environment. In terms of classical agents Wooldridge is perfectly right of course, but the GToC isn't classical, nor is the environment the agent will act in.

A small advantage is gained by the GToC regarding the uncertainty principle of being able to perform an action. By stimulating actuators, the agent sometimes needs to consider whether the action has been performed or not. It can really never be sure that an action has been performed. If it has been found not to have been performed, the agent needs to consider whether it should stimulate more actuators to be emitted. In some cases the problem is more physically identifiable. (A big rock which is too heavy to move may be blocking the road and the agent is never able to move this by himself.) So the mechanism of performing actions now consists of the emitting of actuators and possibly a checking function whether the action has been performed. After every action the agent should thus consider whether it's important enough to check whether it has succeeded and possibly whether it has succeeded.

Because of efficiency matters it's important to realise whether the priority to check whether an action succeeded is high enough to actually do so. If for every action the agent would have to check whether it succeeded this would slow down the agent. For example, as may be noted in for instance human behaviour we often assume things to be true, even though they really aren't. This may then sometimes lead to a failure. For instance when standing up with a *sleeping leg* which sometimes isn't noticed to be asleep this leads to a failure. Normally we might expect the leg to work, so we don't constantly check whether it actually does. In some exceptional cases we don't perceive the leg to be asleep and thus we fall. If we would have to check whether our limbs were working with every action we performed, this would take up way too much time. In terms of computer science this would be like checking every time whether the hard-drive still worked after having written a single bit to it. It just takes way too much time and the action should be assumed to happen properly.

The agent's state which is normally modeled as part of the agent can be perceived by the agent according to the GToC. This means that it's part of the environment. Although the number of inputs of the agent control architecture will now be a bit bigger coming in from the environment, the same mechanisms may be used for the perception of its own state as well as the environment. Its own state is reduced to the environmental state. So the basic interaction mechanisms which change the state of the world may now be seen as in figure 4.1.

As may be seen in figure 4.1 the environment contains actions. These are the physical changes to the environment. The dormants[4] and actuators induce these physical changes. Dormants and actuators may also be processed by other processes. In case of other agents these will not emit dormants, but actuators. The environment itself of course contains dormants. In order for one dormant or actuator to be transferred to another type of dormant or actuator a process is needed. For instance a dormant contained in a particle as energy may be emitted as light, another dormant. On the action of the electron in a particle falling back from a higher energy level to a lower energy level, the process associated with

[4](Hobo, 2004b, definition 5.2)

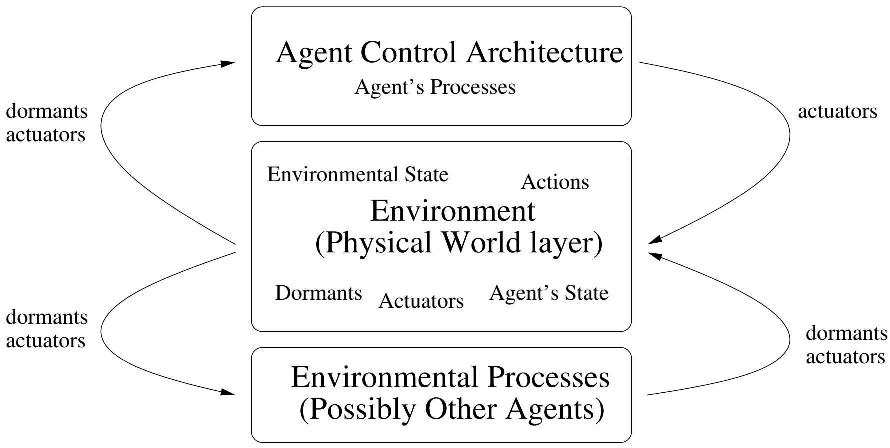

Figure 4.1: The way the agent's world works

doing so will emit the energy in the form of new dormants called light particles into the environment. Now new things may happen to these dormants. Basically an *action* is now just a name for a *change of state* of the environment, with a process *associated* to it. Consider an action as a template-name which still needs to be associated with a process.

4.2 The Agent's Task

An important part of the agent is its actual task. In order to determine its task it's important to realise with what kind of an agent we are dealing here. When this has been made clear it's then possible to determine how this is reflected within the architecture as well as the environment.

Wooldridge makes a clear distinction between *maintenance tasks* and *achievement tasks*. The agent envisioned here is concerned with the maintaining of its own processes as well as child processes. This means in some cases that it will have to set temporary goals to achieve in order to keep the general process going. This clearly illustrates that it isn't possible to clearly define whether the agent is concerned with a maintenance task or an achievement task. Although the global task is a maintenance task the subprocesses containing both kinds of tasks sometimes will need to receive higher priorities. This then serves the global maintenance task but may mean that an agent temporarily has an achievement task as its goal. So an agent should not only give priority to actions, but also to tasks. Based on this priority the agent will need to determine what to do.

The *priority function* for an action or task is highly dependent on present circumstances as well as history. It cannot only be determined by the type of action or task to perform. This means that the priority function is a dynamical function over time. In case of a being both the actions and the task are executed by emitting actuators. When speaking of assigning priority to an action or a task this then of course means assigning priority to the emitting of specific associated actuators.

An agent that performs a task within a certain environment doesn't always recognise that it was performing a task. Quite often the action an agent is engaged in is the only thing recognised by the agent. In such a case the agent needn't be aware of a more global task. Even more often it just does its task and engages in actions without even spending thought on that it does it.

The task itself is a constraint imposed on the agent over time which has molded the architecture of the being to suit its task. This molding doesn't only change its physical architecture, but the control architecture with it. So the control architecture is also largely genetically confined to a certain structure. Of course different parts may be identified as separately identifiable control systems. Each part may take care of a specific task and generates an output based on which further conclusions may be drawn. Some of these control systems have loosely been defined in chapters 2 and 3.

The architecture will for a large part have to be grown using genetic programming. Although the different architectures may be grown separately they will need to be merged into one larger architecture. The parts of the control system which glue all the separate systems together will then have to be evolved in the same way. But can this still be said to be an agent as most people would refer to something to be an agent?

These grown architectures do behave according to basic agent properties. The fact that they are grown doesn't mean that they don't have these properties anymore. The properties will just to some extent be hidden for us to see.

The way these architectures are grown may be directly related to the *utility function*. The utility function stimulates the growth of the architecture in such a way to perform a certain task. What it now basically does is that it measures its performance for a single run of an architecture. When the performance belongs to the top few performances, the process stays. When it doesn't, the process goes.

Basically what I propose is *agent synthesis* as also discussed by Wooldridge. The agent synthesis now becomes a little bit more complex. It has to be done by genetic programming and evolving the right structures (possibly to fit a given physical structure). Of course everything may first be simulated in virtual worlds before possible non-virtually embodied agents are introduced.

4.3 Control Systems

In the previous sections the general architectural constraints and the agent's task have been discussed. What's now still needed is a certain control system which guides the agent through its task towards its goals. What would be the general make-up of such a control system?

A control system, as seen from the outside, has certain inputs which through some kind of manipulation of inputs is turned into outputs. The inputs are formed by sensory data, or information gathered from its senses. These are basically the collected dormants and actuators from the environment. The outputs are formed by the emitted actuators.

Figure 4.2 illustrates how the control system generally works. The signals from the environment in the form of dormants need to be translated into a form processable to the control system. In compliance with the GToC these translated signals are called transmittees.[5] These are then processed by the

[5](Hobo, 2004b, definition 5.7)

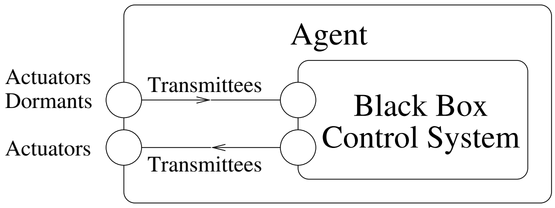

Figure 4.2: The control system of the agent

control system, which finally emits transmittees again. These transmittees are then translated to actuators which may induce actions in the environment.

What's now needed is a system to control all the data gathered from the sensory systems, after which certain outputs may be formed. There are a few basic control systems that should be in place. The control systems can be separated into different kinds. This and the associated architecture is illustrated in figure 4.3.

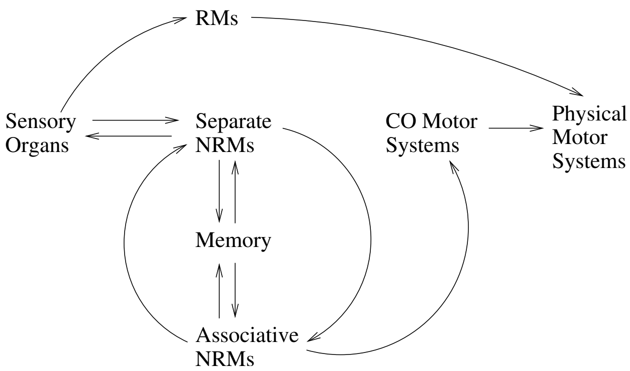

Figure 4.3: The data flow in the control architecture of the agent

The function for each of the control systems may be identified as follows:

1. **RMs.** For each of the four types of senses there should be sensory reflex systems. These are directly associated with stimulation of physical motor systems.[6] Each of the RMs has been described previously for each of the senses in chapter 2.

2. **Separate NRMs.** For each of the four types of senses there should be an associative system which determines the current input and the amount of change in the input. This may then be linked directly to the inputs to increase the focus on certain aspects of the inputs. So the sensory organs form inputs as well as outputs for the separate NRMs.

   Next to the sensory organs the separate NRMs are linked to memory. This memory may influence the NRMs based on previous experience. To do so the separate NRMs receive inputs from the memory system. Just as well the outputs of the separate NRMs are sent to memory.

[6] In our case: muscles.

The current input as well as the change of the input should be sent as an output to the associative NRMs. These in turn may also induce changes in perception by influencing the inputs from sensory organs. This happens through the separate NRMs, which means that the associative NRMs form an input for the separate NRMs.

Each of the separate NRMs have been described previously for each of the senses in chapter 2.

3. **Associative NRMs.** The associative NRMs should determine the relation between each of the senses. This means that the associative NRMs receive their inputs from the separate NRMs. This way the relation between the current inputs as well as the change of the inputs for and between each of the senses may be determined.

   Part of the input is also formed by memory. This on the one hand influences the associations found based on past experience. On the other hand the found associations are sent to memory.

   Finally the associative NRMs are also linked to the motor systems. This may then lead to physical changes within the world.

4. **Memory.** The memory receives its inputs from the both types of NRMs. Its outputs are linked to these NRMs.

   Since memory is largely dependent on representation I would hereby like to propose that the memory is also the representational architecture responsible for our feelings. This is in accordance with the notion that qualia are also induced based on experience[7] because our experience, our memory, now directly induces these qualia. Of course the qualitative system may also be seen as a separate system. But since they are so strictly linked to memory it seems a bad idea to physically separate them. Probably its best to consider the qualitative system a subsystem of memory.

5. **Motor systems.** The CO motor system is used to induce changes into the world. It receives its inputs from the RMs and the associative NRMs. The outputs are directly linked to physical motor systems.

   What should be explicitly noted is that the reflexes are directly passed to the physical motor systems and not to the CO motor systems described here. This to ensure an as short as possible reaction time as possible.

4.4 Final Remarks

The general architecture of an agent with conscious properties can now be divided into two parts. Each of these parts implies for a large part whether a being may perform its conscious tasks within a given world. Not being able to perform the tasks will not mean not being conscious, it just means that the being is being limited in its actions.

On the one hand there's the control system. The control system will have to be adapted to the specific task of a being. To some extent the being may learn to perform properly in its world. For instance when a being is exposed to

[7] Section 2.4

a dark world it may learn to rely more on its auditory and haptic senses. Of course the being doesn't necessarily have all of the senses we have, which will then mean that it will have to be more specialised in one of the other senses. This subject has been more eloquently dealt with in chapter 2.

On the other there's the physical architecture. This physical architecture is largely associated with proper propagation through its world. Here *proper* means that the physical architecture is adapted in such a specific way as to make use of the physical properties of the PW. This has to be done in such a way that the being may move using as optimal mechanisms as possible in relation to its super-imposed task. Chapter 3 discussed the physical architecture to a somewhat larger extent.

Chapter 5

Ethics

Consciousness research regarding the amounts of interest and understanding we are developing for the subject may possibly be claimed to be at its peak. The insights gained through work and research done in different research areas has propelled us in the direction of understanding what consciousness is all about. In view of these developments I wrote *The General Theory of Consciousness (GToC)*[1]. At the beginning of my work I realised that it was very important to not only discuss the theory itself, but also the ethics concerned with such a theory. At first these ethics were captured in the thesis itself. But with the restructuring of the thesis in the final stages of my assignment, the ethics were moved to a part where they couldn't be done any justice. In order to do do ethics justice, the ethics are now featured in this chapter. By coupling them to agents they have now also been placed within their proper context.

The generalisation of being as proposed by the GToC of course leads to a further generalisation of laws and rights. When looking at law as we have it in our society, it almost always applies to humans only. The rare occasion where a law is introduced for specific animals is considered quire. By generalising the principle of being however, determining relations beyond social structures alone, it should now also be possible to introduce law without naming specific beings. It should now be possible to introduce law based on relations found between different beings, considering also their environments. These may now also be directly related to the GToC by looking at social networks and the way they work.

Law is one of the most direct ways of expressing ethical rules, so I would like to introduce laws permitting beings to live their life the way they want to. But how can this generalisation be named without naming specific beings? This will mostly have to happen based on natural organisations in structures as they have for instance been more generally determined in biology. We should then also speak about rights in terms of these structures.

When non-organic beings would come into existence, they would have to have certain rights also. But these should be determined not by what they are, or how they look, but by how they behave. Is it possible to derive such laws? Should these beings be brought into existence to start with?

We from an ethical point of view regarding rights and duties shouldn't make

[1](Hobo, 2004b)

a distinction between a black, pink, yellow or any other skin colour, just as well as race or sex. But, to put it bluntly, why should we then make a difference between organic skin and metal?

Here follow generalisation principles of law and their consequences. These are arranged in a few sections. Section 5.1 describes the application of law. Section 5.2 describes what should be considered a crime and what not. Section 5.3 describes how the verdict should be passed based on the crime. Section 5.4 describes what determines the social context which may be disturbed by a crime. Section 5.5 establishes whether we are allowed to basically play God and create beings.

5.1 Application of Law

One of the main problems with law is in what manner which laws should apply to which beings without discriminating beings. To illustrate this I pose five examples.

1. When a human is swimming in shark-invested waters, I wouldn't call a shark a killer because he ate the human swimming in the waters. The human entered a domain he could have easily avoided and was responsible for assuming position at the bottom of the food-chain. Even though a human could claim to have no prior knowledge, he's still outside of his own habitat.

2. When a lion enters a city, this is also the mistake of mankind. Mankind itself has either accidently set the lion free from for instance a circus or built the city in the habitat of the lion. Most people would still immediately try to kill the lion. The lion can however just as easily be put to sleep temporarily and taken somewhere else. It's not the mistake of the lion when it eats someone, since it is in his nature to feed when hungry.

3. When a lion or a shark attacks humans, humans do have the right to defend themselves as they would in nature. The technology used to defend ourselves sprung forth from nature's ingenuity. So when we carry a weapon we may use it. When the lion doesn't do anything, we don't have the right to kill it, because it is not longer in our explicit nature. Food is provided by separate organisations within society occupied with food. People with different tasks thus shouldn't bother with food even though there is no specific reason to do so and there are plenty of other options.

4. What should be noticed is that a human provoking a lion is the attacker and not the attacked. The human which provoked a lion to attack then is guilty. Just as well a human provoking a shark by entering its domain with for instance a harpoon quite clearly has the intention of killing it. The human could have easily avoided the hazardous action by not swimming in the waters. The swimming doesn't weigh up against the death of the shark.

5. In some cases beings adapt and form social bonds with other beings. For instance wild dogs and wolves when being fed by humans were domesticated and evolved into the dogs we now have as pets. They are trained

and educated by their boss and learn to behave. When they don't they should be set free in a special nature reserve because they conform too much to a more natural habitat and type of behaviour. Dogs which have been previously domesticated should of course not be bred to become brutal killers. This robs them from their social context in living together with all of society.

What now needs to be determined is how habitat and social restrictions can be caught under simple rules. As previously stated for instance dogs have been domesticated. Sometimes however they tend to return to their previous nature. As it is now, when a dog attacks a human being for no apparent reason, he is put to sleep. Instead I propose that he should be set free in a nature reserve. It remains in his nature, as it is in ours, to hunt for food and prove himself the strongest.

statement 5.1 (applicability of law) *All law is applicable to beings with a sense of ethics which is portrayed by them fitting themselves into social structures of society.*

statement 5.2 (equality of being) *All beings are considered equal and all laws apply in such a way to beings that there's no physical or non-physical discrimination based on exterior properties as well as heritage.*

When a being's nature is physically expressed, a being should be granted a habitat where all beings believe this to be so. The social context in which a being is placed should be specific to the being.

statement 5.3 (right of own natural habitat) *All beings have the right to live in the natural habitat to which they are physically adapted.*

statement 5.4 (expected conformation to rules) *When beings are removed from their natural habitat and are robbed from their social context they cannot be expected to conform to the rules of the social context into which they are placed.*

Every being has a right to decide for himself. The being itself may of course when possible say what is his social context. For instance when one human calls another human a fish, the other human cannot just be thrown into shark-invested waters.

statement 5.5 (freedom of self-expression) *Every being has the right to say what he perceives to be:*

1. *his own position and*
2. *the social structure of the world within perceivable reach.*

Sometimes beings in some societies are withheld information which greatly influences their lives. For instance when some factory injects all kinds of toxic goods into the environment, the surrounding beings should know in advance. This is the only way it may be stopped beforehand or the beings may choose to move away.

statement 5.6 (right of knowledge) *When certain structures physically influence a being's social structure even though he doesn't know of these structures, he has the right to know.*

Just as well every being has the right to defend himself. So when a dog is or becomes a vicious animal, we are allowed to defend ourselves and in defense quite often kill it when attacked. But when possible the dog should be returned to its natural habitat.

statement 5.7 (right to defend oneself) *A being when attacked in any type of social context has the right to defend itself.*

The law should prohibit the condemning of a crime for the sake of the having of a condemned itself. It occasionally happens that someone has to suffer without reason and without guilt. This should never be the case.

statement 5.8 (circumstantial evidence) *A being should never be condemned based on circumstantial evidence without physical proof.*

It doesn't matter what type of being a being is. It only matters why, when and how a crime was committed. When done with or without proper reason, when committed by someone from a higher or a lower social class, when committed against a higher or a lower class, it doesn't matter. Independent of social class the verdict should remain the same.

statement 5.9 (application of verdict) *All law should be applicable to any being where for all beings should follow the same sentence based on the same behaviour given the same circumstances.*

statement 5.10 (defining circumstances) *Circumstances are defined by the relations which hold within a moment in time corresponding to certain behaviour, where they are non-dependent on heritage or exterior make-up of the being.*

5.2 Committing the Crime

Now suppose that a being supposedly committed a crime. How should be determined what a crime is? And when determined what the crime is, what is then the sentence that goes along with it?

definition 5.11 (crime) *A crime committed by a being **A** is the disturbance of some other being **B**'s social structure in the following manner:*

- *A being **A**'s actions adhere to one or more of the following items:*
    1. *The disturbance of **B**'s social structure has a negatively perceived effect by **B** or negatively influences the continuation of the process of **B**.*
    2. *A different choice was available to **A** and would have prevented a negatively perceived effect by **B** or the negatively influencing of the continuation of the process of **B**.*
    3. ***A** didn't act with the reason that any other choice of action would have led to the cessation of his own process.*
    4. ***A**'s choice of action on the continuation of processes when chosen to do so together with **B** didn't coincide with the choice made by **B**.*

- ***A** was aware of the consequences of the above actions when preparing or executing them.*

Although these items seem to be quite clear they have some far reaching consequences. These consequences implicitly follow from what has been defined. When not committing a crime all the defined general laws should hold.

One consequence is that every being has a right of privacy and a place of its own. This is needed for a being to live his life the way he wants. This on the one hand guarantees the being's own social structure. On the other it makes sure that the being can live his life without disturbing another being's social structure.

Another consequence is that a being may not kill but does have the right to defend its life. Defense may be associated with the possibility of the cessation of the other being's process. Here the being shouldn't have had any other way of defending himself where the other being would have survived. If a being can defend himself properly without killing the attacker he should do this. But this ability of course also differs per being.

A final consequence named here is that a being may not rape another being. When a being is raped it is without this being's consent that intercourse takes place. Raping basically equals the disturbance of the raped being's social structure and it may also mean the continuation of processes without consent.

The only way these laws can then be falsified is because the falsification of either one of these laws is inherent within the whole of the social structure. For instance male ducks tend to rape female ducks. Or a barracuda eats smaller fish. Human society however is built with the purpose of living together in peace. Anything that goes against peace is thus perceived as a crime within human society. But then again, we are physically built in a way that we like to eat meat. So we do kill animals which are below us in the food chain. Still passing a verdict should be made as easy as possible. Otherwise discussions over smaller crimes may hold up discussions that matter more.

5.3 Passing the Verdict

When it is determined that a certain crime has taken place it's still a matter of opinion what the punishment should be. Even within specific types of crimes different sentences may be passed based on circumstances. For instance a being may be an accomplice or the physical executer of the crime. An extreme case would be where the accomplice actually was the one coming up with the plan and the executor being one who wasn't aware of its actions.

The most general guideline is that the punishment should reflect the crime. Basically the tougher the crime the tougher the punishment. But does *an eye for an eye, a tooth for a tooth* count as a good guideline to executing law?

Errors in judgement should as far as possible be repairable, so judgement beyond repair may not take place. It still happens that people who have been found guilty should be set free because they hadn't actually done anything. In such a case it should remain possible for always to set people free.

Just as well the person executing the sentence shouldn't be reduced to a criminal himself. This last part is of course also reflected in the proper reason in executing an action which would otherwise have been a crime. So a person executing a punishment cannot be held responsible for specific executions. He can however be held responsible for voluntarily taking up the job to do so.

definition 5.12 (severity of crime) *The severity of crime is the amount of disturbance caused in the disturbed social structure.*

statement 5.13 (severity of sentence) *The sentence should be as severe as the crime with the possibility of reversing the sentence to the largest possible extent in case of errors in the judicial system.*

Of course there should also be a guideline for the amount of freedom which should be granted to beings which haven't committed a crime. The larger the amount of consciousness the larger the amount of space they will need to live in.

Just as well when they perform a very specific role in the survival of the whole world they should never be put in a cage. The only exception may then be that there are too many of them. This may lead to problems regarding the survival of their own as well as other species.

statement 5.14 (freedom of being) *The minimum amount of freedom a being which hasn't committed a crime should receive should be dependent on the maximum of the two direct relations to:*

1. *its amount of consciousness and*
2. *its role as a species in the whole of the world for the continuation of the whole of all beings.*

The amount of consciousness is a hard thing to measure. At first when looking at beings with higher levels of consciousness these beings will organise themselves in social structures to survive. But when looking at even higher levels of consciousness, some beings may choose not to adhere to these social structures because they don't think they are right. Others will just accept them for what they are under the assumption that they cannot be changed. These beings will fit themselves into social structures just as much as needed to survive. There are yet many other variations, all highly influenced by personality. Personality is partly genetically defined and partly formed by experience. I've tried to draw an estimate graph of beings fitting themselves into social structures based on amounts of consciousness. This graph, which is portrayed in figure 5.1, should be considered as an *average* estimation. So there are basically small variations regarding social adaptation for each level of consciousness.

You may now wonder how this graph should be interpreted. When looking at nature the most organised of beings are basically insects like ants. So the amount of consciousness of an ant is linked to the highest possible social adaptation level which is shown in figure 5.1. We humans are far less organised and thus less socially adapted. We do behave according to hierarchies, but in an individualistic kind of way.

What all this basically means is that a being should be given the liberty to take up its role in nature as provided to him by evolution in a biological sense. Non-biologically verifiable goals should thus be abandoned to a limited extent. Beings within social structures sometimes occupy biologically non-verifiable tasks. These tasks however may have more far reaching effects which

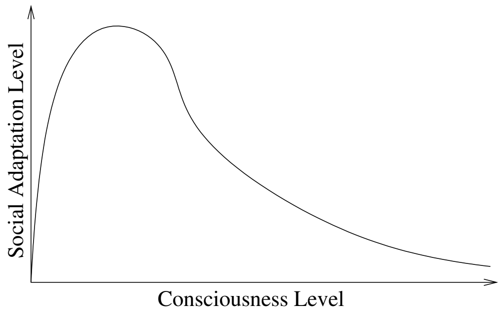

Figure 5.1: Social adaptation and consciousness

are biologically verifiable. But how does this reflect on the current beings as found in nature?

To start with us, we can for instance to a relative extent still eat meat. According to our genetic make-up and our build we cannot avoid eating meat without becoming unhealthy. The way it should be can be seen in for instance the cultural structure of the native Americans before the conquerors set foot on their country. They used to live in harmony with the beings found in the country, not killing more than necessary and making use of all of the material as found on the killed animals.

Basically all predatorial roles as initially found in nature not determined by cultural blood-lust should be tolerated. For instance sharks within their own living environments have the right to eat meat. If we as humans enter their environment they have the right to eat us.

What changes for us in the present is the amount of meat people eat. There's no need for mass consumption, so it is a crime to eat more meat than is necessary. So basically in percentages we should be bound by law to eat the relative amounts of food as dictated by the usual health-charts.

Although this can hardly or actually not be controlled supermarkets can be forced to change their assortment in such a way to provide healthy diets. It isn't possible to control people's diets, because some people have to eat more meat than others (or just more food in general). So the basic idea is that we shouldn't kill more than we actually need.

There are of course a few details. People with medical problems sometimes grow fat not by their own or the supermarkets choice. There is nothing they can do about this. Just as well people have to eat relative amounts. So one person may eat more meat, but this would then also indicate that this person would have to eat more vegetables and other necessary stuff. The only thing that should be approximately right would be the percentages of food eaten relative to each piece of the meal. But even these percentages change from person to person.

This was just an example. It does strikingly show how some of our abundant behaviour should be controlled to limited extents. The only problem is that we only see problems when they have gotten out of hand. So what should we do then?

For any other type of abundance which influences nature in a large negative way follows that it should be prevented. The only way things like this can be controlled is by controlling it at the source of the problem. This prevents any way of starting and spreading the hazardous conduct.

statement 5.15 (responsibility) *The direct effector of the induction of negative effects should be held responsible together with those who willingly and knowingly cooperate in effecting the induction of negative effects.*

5.4 Social Context

The social context into which a being is placed should be dependent on the nature of such a being. Within different species of beings within different habitats their nature becomes quite clear. Their nature reflects their tasks which nature has given them. Most unfortunately humans have grown a bit beyond their tasks, which is illustrated by the fact that we come up with goals of our own. In case of most other species this isn't so. Who has ever heard of an artistic lizard? This makes things a whole lot more difficult in determining the social context of our being. On the other hand in case of other beings they may just do their tasks. That's why they're here and we can't blame them. But how do we now determine the social context of a being?

statement 5.16 (social context and goals) *The social context of a being is dependent on the goal which a being has chosen to pursue in his life.*

This goal may in terms of beings with lesser amounts of consciousness[2] be determined by nature. Some beings have chosen actively to pursue other goals in life. This differs per being based on the nature of this being.

The only goal a being may not set his mind to is absolutely doing nothing without contributing the least to society. A being should do at least as much as a being would in nature to preserve itself in nature. If a being doesn't he disturbs other beings social structures by profiting without returning anything. A being has not only the moral but also the physical obligation to contribute something to society.

statement 5.17 (contribution to social structures) *The physical contribution of a being to its social structure should at least be equivalent in amount to what a being needs to do in nature to survive.*

A special debate could be formed around the principle of relaxation or spare time. Relaxation is also needed for the continuation of processes. Different forms of relaxation should be allowed to be provided as long as they are provided in a separate social structure. For instance a painting should only be put up in places with the consent of the inhabitant of the social context. Just as well when prostitution is performed willingly by the prostitutes it may be allowed also within certain boundaries.

The boundaries of one social structure should of course not disturb any other social structure. Here the knowledge of the existence of a social structure doesn't mean that it disrupts another social structure. It should in some physically perceivable way be disturbing a social structure before it can be counted as a crime.

[2]Not lesser consciousness!

statement 5.18 (inflicted damage by knowledge) *When knowledge inflicts damage to a being and no physical damage is done in the real world, the being inflicts the damage upon itself.*

definition 5.19 (physical damage) *Physical damage is a physically perceived crime by and to that same being.*

Something else that should be stated is that different social contexts should have their place in society. There is basically enough room for everyone. People who want to go to some religious housing can go to such religious housings. People who want to go to a bar or maybe a tea-house can go there. When they are there they should behave appropriately.

But what should now also be stated that in going somewhere there may be no discrimination based on people's exterior properties. For some occasion you might want to only allow people wearing pants. The pants aren't part of a being's exterior properties. But what may not be stated is that only men or women or blacks or whites should be allowed.

The rules should be stated thus that it's easy for beings to adapt to such a social context. In some cases the way beings want to go somewhere can be free for the beings to decide. These places should also exist. There basically should be many places for any kind of being anywhere in the world.

Routes should be provided in a sufficient manner to reach one place from the other without having to disturb any ones social structure. This means that someone going to a coffee-shop could be provided a place nearby to sober-up again. If this isn't possible this person should be granted proper passage home without disturbing anyone or being disturbed.

statement 5.20 (separation of social structures) *Different structurally conflicting social structures should be physically separated to make way for each other.*

One mistake which is often made is that people claim that there's one proper social structure to which all beings should adhere. This quite clearly isn't inherent to the nature of the whole of mankind. With the amounts of consciousness we have received we have grown from a purely conformist nature into at least more general conformist natures. Sometimes even some people refrain from any kind of conformism which can usually only be shown by their opinions and not by their external make-up.

Many social contexts come forth from religious ideas or other ideas regarding for instance the perfection of mankind. People are of course allowed to have these ideas. People cannot be not allowed to think for themselves and perceive the world in their own way.

What should be noted is in which manner these ideas are maintained. If people have these ideas for themselves and ways to maintain information about their ideas this is allowed. But they shouldn't be allowed to force their ideas onto others in any kind of way.

People should turn to ideas voluntarily because it adheres to their social context, structure and beliefs. Offering someone relief based on some religious or other social structural idea shouldn't be allowed. If someone however thinks that there's more out or in there and finds some religion or other expression of this feeling then there's nothing wrong with grouping with like-hearted people.

statement 5.21 (social beliefs) *Instead of forcing them on beings, social beliefs should be formed by beings themselves and beings should turn to those beliefs because of their own conviction.*

5.5 The Principle of Creation

This thesis is concerned with the GToC, which consists of basic architectures and functionality which should be implemented. Suppose that we would for a change learn from the past? Now we would realise that with these possibilities arise certain responsibilities. What are these responsibilities and can we live up to them? May we consider ourselves to be suitable creators?

When coming into being nature has provided us with a place in the world and a felt sense of purpose. This sense of purpose comes forth from evolutionary principles. This sense of purpose basically ensures the continuation of processes. Is it possible to create a being? Can we identify a sense of purpose which may be associated with such a being?

statement 5.22 (right of purpose) *Every being has the right to have a naturally induced sense of purpose.*

To start of with we can of course create beings. We've been able to as a species since our ancestral species first came into being as single cells which split in two to create two new offspring. These beings have, just as we ourselves do, a place in the world. They have social or practical relations into which they fit. This as stated earlier is preprogrammed by evolution. But when a being is created by us, what reason to exist does it have other than because we built it?

An artificial being built by us didn't spring forth from evolution. It sprung forth from our ingenuity. It doesn't naturally ensure the continuation of any, not even our own, process. So it is in no way related to evolution. Evolution is what purpose makes. Us creating beings would mean creating for instance kamikaze-pilots which have been *taught* not to think for themselves. This isn't accepted within our culture, so why should any being that sprung forth from the ingenuity of our culture possess such motives? Or when we create beings to fit into our culture, they would need to be suited to us in every way. So there's basically no way a being may be created other than to create it in our own image. This can only be established by natural cause.

statement 5.23 (creation) *Beings may only be brought into being through natural cause.*

To what extent can we create then? We should be conscious of the fact that it's completely unethical to create beings in other ways than the ones handed to us by nature. What we create should be limited to simple *input state* $\leftrightarrow$ *action output* systems where such associations between *input state* and *output action* may be learned. But the *input states* as well as the *output actions* should be predefined limited domains. So basically the thing created by us by other than natural mechanisms should be limited to predefined domains with clear boundaries.

statement 5.24 (limitations of creation) *Mechanisms created by us should be limited to:*

1. *simple* ***input state*** $\leftrightarrow$ ***action output*** *systems where*

2. *such associations between* ***input state*** *and* ***output action*** *may be learned and*

3. *the* ***input state****s as well as the* ***output action****s should be predefined limited domains.*

Conclusions and Recommendations

In order to properly describe an agent different goals had to be met. For each of these goals now follows a discussion whether it has actually been met.

For each of the senses different subsystems and their relations have been established. Each of the senses has been associated with certain RMs and NRMs. These have then been properly related to the agent's physical mechanisms.

The influence of the physical mechanisms has been globally identified, abstracting for any type of sensor the basic demands it should meet. This way the embodiment of a being and its influence has been properly defined.

In order to have a proper control architecture the relations between all of the mechanisms has been defined on a more abstract level. The senses may now be fitted into this abstraction of the control architecture of an agent. Together they then form a proper control system for the whole of the agent.

In order not to break through any ethical boundaries into unethical domains, a proper ethical foundation had to be constructed. This has been done by constructing a general specification of how beings should fit into social structures.

In conclusion, every goal which had been specified has actually been met. This then leads to a general architecture for an agent which can easily be customised to fit into its specified requirements. This has been done by separating each of the parts of the design into different abstraction levels. So the global architecture which has for instance been constructed in chapter 4 remains complete when one of the senses isn't instantiated.

In the future each of these designs will have to be tested piece by piece in order to determine how they may be instantiated. This should be done keeping in mind the ethical boundaries. When each of the separate components have been constructed they may then be fitted into the general architecture of the agent. The physical architecture should adhere to the properties of the environment the agent is placed in.

Bibliography

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

P. Eggenberger Hotz, G. Gómez, and R. Pfeifer. Evolving the morphology of a neural network for controlling a foveating retina - and its test on a real robot. In M.A. Bedau R.K. Standish and H.A. Abbass, editors, *Artificial Life VIII: Proceedings of the 8th International Conference on the Simulation and Synthesis of Living Systems, Sydney, Australia*, pages 243–251, 2002.

Emile Michel Hobo. (appendix ii) derivative ideas and considerations based on the general theory of consciousness. 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.

Akiya Kamimura, Satoshi Murata, Eiichi Yoshida, Haruhisa Kurokawa, Kohji Tomita, and Shigeru Kokaji. Self-reconfigurable modular robot - experiments on reconfiguration and locomotion. In *Proceedings of 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2001)*, pages 590–597, 2001.

M. Lungarella and L. Berthouze. Adaptivity via alternate freeing and freezing of degrees of freedom, 2002.

Mihoko Otake. Gel robot online. http://www.jsk.t.u-tokyo.ac.jp/ mike/gelrobot/index.html, 2004.

C. Paul and J.C. Bongard. The road less travelled: Morphology in the optimization of biped robot locomotion. In *Proceedings of The IEEE/RSJ International Conference on Intelligent Robots and Systems , Maui, Hawaii, USA*, 2001.

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

G. Sandini, F. Panerai, and F.A. Miles. The role of inertial and visual mechanisms in the stabilization of gaze in natural and artificial systems. In J.M. Zanker and J. Zeil, editors, *Motion Vision - Computational, Neural and Ecological Constraints.* Springer Verlag, 2000.

Tokyo Lectures. Aritificial intelligence lectures in tokyo. http://www.tokyolectures.org/, 2003. Lecture 3.

M. Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley & Sons Ltd., 2002.