# The General Theory of Consciousness

E.M. Hobo (MSc) / Æmilius — *14th of January 2005*

## Acknowledgements

What I here wish to acknowledge is the following. The first theory as I presented it was a minimal model, and was correct in the way it worked. In order for people to work with it, however, I've had to change certain names of layers, and I even added new definitions. These definitions should clarify the theory a bit further, and make it a bit more comprehensible to the reader. The old definitions remain largely intact, incorporating certain principles as introduced by the new definitions. I hope to have clarified the principle of discussing consciousness in more general terms a bit further.

## Introduction

Many people are intrigued by the concept of consciousness. There are many different sentiments and explanations around that try to explain it. No matter what they are: let's call these "the theories of consciousness". These theories most of the time don't have too much in common. Quite often the concept of consciousness even changes from theory to theory. Consciousness is thus large that it can't ever be captured in a single statement without being left with more questions. Perhaps it's a better idea to propose a theory of consciousness without defining consciousness itself. Maybe even, can this theory capture or relate all theories of consciousness? What now would such a theory look like?

There are so many concepts related to consciousness. One person would describe consciousness as the ego. The next would say that true awareness is the tool to kill the ego. The only thing different theories have in common is the fact that we come to conclusions. Whether this conclusion is spiritual freedom or experiencing feelings shouldn't matter in the proposed relational processing structure. The processing relations are the only things that just are in any type of consciousness. Or perhaps I shouldn't say *just are*, but I should say that they are given meaning through different theories of consciousness. But can't we relate different types of consciousness by capturing these processing relations?

I believe we can by describing how certain input from a reference frame may be related to certain output into such a reference frame. By describing the processes on top of such a reference frame, we may then describe the building blocks that make beings be. A large amount of these building blocks working together will then lead to consciousness itself. You may wonder how many of these building blocks are needed to have consciousness. Then you may start to wonder whether it was such a good idea to use this method to begin with.

The reason why describing basic building blocks suffices for gaining a far better understanding of consciousness is the following. No matter how large the process of consciousness may grow, it will always try to behave according to the basic properties of the building blocks it's made up of. Although the complexity may sometimes lead to radically different behaviour, these are extreme cases, and even these can be better understood by understanding the origin of consciousness. Consciousness originates from its building blocks, and these are ingrained in its essence.

## 1. Arranging the basic processes

Since one process depends on the other, and all originates from and ends in a reference frame, the theory proposes to organise the processes in layers. The bottom layer will then be the *reference* frame, which can act as the soil from which all other processing stems. But suppose this soil produces certain qualities which may be considered valid inputs? Suppose this soil is also receptive to outputs, which behave according to these same qualities? How can processes then make use of these qualities?

These *qualities* will need some kind of messengers where incoming and outgoing qualities may be posted for *transmission*. This will have to happen in a layer which acts as an interface between the reference frame, and the processing facilities. This will interpret the qualities in an abstract way. By creating an abstract interpretation it's possible to derive conclusions from all the incoming qualities. These conclusions will be made known in the reference frame as well. This means that the conclusions have to be interpreted for the reference frame, to be able to pass them to the reference frame itself.

In order to come to a conclusion, there has to be a beginning, and an end. Between the beginning and the end lays a path which takes the beginning to the end. This path is organised in different ways, and may follow different routes, of which some more, and some less optimal. What you may now imagine is a map containing vast *networks* that may contain one or more of these solutions found by consciousness. This map may then be used to indicate the path of reason leading to certain conclusions.

A problem then is that conclusions change with times. So for every path there needs to be a clear frame of reference also associated with time. Space and time locations define circumstance, and only from there can a conscious being reason. So when a mapping is created we have space. If we point to space, and add a time we have logical reference. Let's call this space and time pin-pointing the pin-pointing of *locality*, which then defines time, and space.

Of course when many paths of reason can be chosen it's necessary to take a pick. Which ways of reasoning are chosen and which aren't? What should be taken into account and what should be ignored? What should weigh in more heavily and what should be taken lightly? It's all a matter of organising, selecting, or more basically

assigning *priorities*. These priorities will also be distributed in the drawn conclusion, organising conclusions in more and less highly prioritised circumstances to achieve. This way different steps in for instance a solution may be taken by prioritising the step which needs to be taken based on locality which also implicitly captures previous steps. So assigning priorities is a very important thing to do.

The found qualities and their priorities now have to be assembled in a report. The report will be a *representation* of the incoming qualities and the way they have to be handled. This has to be delivered to the process that manages all incoming and outgoing qualities for the reference frame. The managing process does so based on a dynamical knowledge base. This processing based on knowledge is here called *cognition*. The process itself will thus be called the cognitive process.

Before handing the representation to the cognitive process it needs to be organised in ways understandable to the cognitive process. In order to be able to make a valuable judgement it's important that the representation is coherently organised and as complete as possible. When the cognitive process draws its conclusions these are added to the next representations. They are also cast back to the reference frame to be made known in the form of qualities common to the reference frame. Because the conclusions will steer the being, but they cannot always guarantee the correct result, they are called *desires*.

The cognitive process basically indicates which direction to go based on the network map. It does so by indicating the direction and possibly also by trying to block all the paths which shouldn't be followed. This in order to make sure that only proper direction may be followed. Of course the cognitive process will not always succeed in doing so and improper direction is then assumed. The cognitive process can then still try to get back on track, either by going back or taking an alternative newer shortest route. The possibilities the cognitive process has to reach a certain or perhaps no conclusion are limitless. There are many unexpected circumstances that may arise.

So the processes are organised in seven layers. These layers are organised in a stack like in figure 1. What I've tried to show is that during the whole process the scope is narrowed down more and more in order to reach a conclusion. It's impossible to judge the whole by looking at the whole. Conscious beings can only make judgement by focusing on what's important. If conscious beings are unable to do that they will be unable to make judgement, which will lead to the failing of consciousness.

This basically concludes the processes themselves. What you should also realise is that management is also based on or perhaps disturbed by sentiments. This, next to more physically representable qualities, would then mean managing based on qualitative experience. So qualities of the reference frame are then physical qualities as well as qualitative experience or, to use a more common term, qualia.

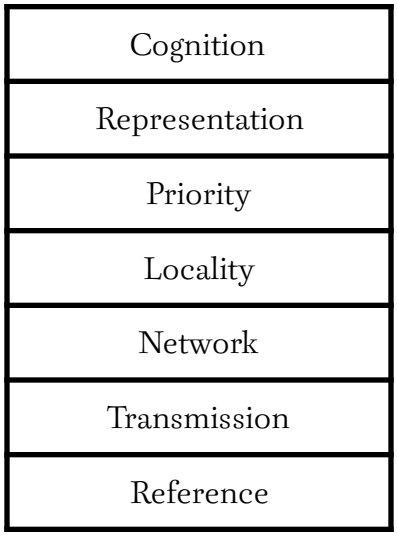


*Figure 1: The seven layers process stack a.k.a. Consciousness Reference Model (CRM)*

## 2. The actual theory

The previously described layered model may be seen as a strange kind of marble. It sends out certain qualities in containers. Throw a whole lot of marbles in a bucket and who knows which marble may communicate with which? This way vast communication networks evolve, that will grow general opinions, solutions, representations, reflections, and much more.

A lot of time is spent on the argument of what kind of qualities there are. There are physical qualities which we can measure, but we also experience. Based on these experiences, with which I would like to address the problem of qualia, we also make decisions. And these experiences change based on our physical perceptibility, and knowledgeable experience during our life-time. They are largely influenced by for instance hormones, and knowledge that certain things are bad for us. How can we discuss these qualia?

In order to discuss qualia there needs to be a structure that can contain the information they represent. Just as well, beings should still be able to argue by numbers.

Beings should thus be able to come to decisions based on knowledgeable experience as well as qualitative experience. Some beings use only either one of the two. Most generally use a combination of both depending on the situation. So what we need in any case is an abstract container that may be used to represent either one of the two, or a combination of both. This container is more generally called a quality.

> **definition [quality]** A quality is a container that may be treated as a signal containing any type of information in any type of representation.

A large part of the theory depends on qualities coming from one point, arriving at the next. But who can guarantee the safety of the messenger? It's not possible to know when or where a message will arrive. Just as well, it's impossible to know where and when a quality will arrive. The only thing that's possible is to make a best estimate. Some estimates will turn out to be more reliable than others, but even the most reliable estimate will not guarantee the outcome. There's always this minor possibility which may prevent the arrival of a quality. Nothing of any kind can be said for sure, except the fact that something happened itself. But this shouldn't be mistaken for our perception of what happened.

> **assumption [knowledge]** It isn't possible to know everything that happens in a world, so you can never be sure which quality is going to reach something in the far or near future, and which not.

What is perceived is largely due to circumstance. For an important part perception is dependent on the arriving of qualities at locations where they may be processed. Because there are no definite states of all that is, that may be reachable time and time again, I would here like to speak not of states but of circumstances. This because I wish to discuss the system as a whole, and not just part of the system. So what I would now like to do is describe circumstances of qualities, which all qualities may go through at some point in time. This doesn't mean that all circumstances will be met at some point in time by all qualities. The possibility just exists.

> **definition [dormant]** A dormant is a quality which has the possibility to in the future hit a receptive and raise that receptive's charge.

Before a quality can reach a point, it has to travel to such a point. During the travel it really doesn't do anything. This is meant by "dormant". A dormant has a wandering property, which means that other than space and time, it has nothing else to relate to when being a dormant. There are however receptives, that are suited to absorb the dormant, and thus its contents. For certain contents to be noticed they should be expressed urgently enough. This means that each dormant may hit a receptive, but enough dormants have to stimulate the receptive in order for their contents to be noticed. How many dormants are needed highly depends on the receptive. Sometimes one will do, but for other receptives perhaps many are needed.

> **definition [activator]** A dormant becomes an activator the moment it hits the receptive and raises the receptive's charge.

Beings are not the only things that may absorb certain dormants in order to process them. Quite often there are many receptives in a reference-domain that also absorb dormants, but don't really do anything with them. So it's important to make a clear distinction between the receptives that are a part of a single being, and the ones that clearly aren't, like a rock.

**definition [receptor]** A receptor is a receptive that constitutes part of the physical relation that is a single being.

So what we may now derive from the above is that it's important to realise that perceptions can only be built based on qualities beings have at their disposal. When qualities don't arrive, for instance because a being closes its eyes, it's not possible to perceive certain circumstances. The being that isn't closed off to certain circumstances may perceive them as its own.

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

As made known previously, a receptive needs to absorb enough dormants in order to get the contents through. So a certain dormant-push is needed to actually let the being make its circumstances its own. If the delivered push is high enough, the receptive will fall back to a (near) zero charge. The built up charge has to be released. This means that new qualities are created.

When we speak of receptors, this means that the receptor passes on the information to the being's processes. Otherwise the information may be emitted as the same or perhaps other dormants into the frame of reference. This may be associated with for instance light heating up the earth. Not all receptives are receptors.

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

When qualities are passed within a being, we can no longer speak of dormants. The qualities are no longer part of the reference frame. This means that a being may now make the circumstances its own, by transmitting these qualities to its processes. These qualities are then captured in what may be seen as a package, defining certain contents. This package is called a transmittee because of its properties.

**definition [transmittee]** A transmittee is a collection of qualities and has been emitted by a receptor onto the network.

These transmittees can only be passed to or between certain processes if communication routes are present. The available routes can be mapped by collecting them in a network. This network may then be used to send transmittees from one point to the next within a being.

**definition [network]** A network is a set of directly or indirectly connected processes using an agreed upon kind of transmittees, to send the transmittees over the shortest possible route.

In order to perceive within one's own time it's as important to absorb the dormants at a specific location as it is to associate them with time. Time doesn't always have to be explicitly modelled, but this doesn't change a thing about the importance of the concept of time. Most relations within a being's world are dependent on it.

> **definition [locality]** The locality of an event is defined by the space-time co-ordinates of that event.

At first the number of qualities might seem to be a good idea to use as the definition of the heaviness of certain incoming contents. Again it's important to consider the factor time. A charge may be built up over longer periods of time by a very light signal. So an approximate of the heaviness of incoming content is the number of transmittees emitted during a certain time-frame.

Since a transmittee may contain a collection of a certain kind of quality, and the heaviness is actually defined by the push of dormants external to the being, it's only an approximate. The heaviness of the actual content as it is, would be the dormant-push. But since the theory is concerned with what's perceived and what not, the heaviness that's perceived must be related to the frequency with which transmittees are emitted.

> **definition [heaviness of content]** The heaviness of content is the frequency with which a receptor emits transmittees containing that content.

In order to come to valid decisions it's important to have all priorities straight, as annoying as this may sometimes be. Based on the priority given to certain kinds of qualities, in combination with the heaviness of content, it's important to combine these two into a potential. This potential is the combination of what some people would call the weight attributed to certain content as well as the amounts of weight delivered. For each kind of quality it's now possible to derive this potential.

> **definition [content potential]** The content potential is a function of the heaviness of content and the expressed desire to focus on that information.

Based on the content potential for each kind of quality, putting them in relation to each other, it's now possible to draw conclusions or perhaps take steps towards conclusions. Putting them in relation to each other means organising them and putting them next to each other. In doing so a wave will be formed containing all potentials. This wave may be used to identify peaks, depths and changes in time.

> **definition [potential wave]** The potential wave is a continuous approximation-function of the collection of all content-potentials according to their contained content.

Sometimes changes arise more suddenly than otherwise. In case of people this may lead to a scare, perhaps followed by relief. In any case it claims attention of the incoming content.

> **statement [unexpectedness of a transmittee]** Unexpectedness of a transmittee claims attention of that transmittee, where not expecting comes forth from expecting the complement of certain received transmittees.

Although unexpected contents are presented on a daily basis, generally a being will strife for a more stable focus on things. In focussing on one specific set of contents, it will need to repress the focus on other contents.

> **statement [focusing on content]** Actively focusing on certain content means actively diminishing the focus on all other content.

The incoming contents as well as the associated focus can now be captured in a certain representation suitable to further processing. This representation needs to capture a certain reference frame associated with a certain time. It will also have to capture the dynamics, the changing of the reference frame in relation to previous times. (A definition of "desires" follows up ahead.)

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

Since it's also important to consider previous times, the newness of certain contents, of certain circumstances is something that's highly influential on perception. If something is new there's nothing to relate to. The newness of course wears off when things become old.

> **definition [newness]** The newness is a function that's reversely proportional to time.

If something is still new the amount of learning that will take place will be relatively large in relation to when something isn't new anymore. There are a few exceptions. A few conditions have to be met in order for a being to learn.

> **statement [conditions of learning]** Learning can only arise when the following three conditions are met (Manzotti, 2003a):
> 1. The being is not familiar with the encountered, i.e. the encountered has a high newness to it.
> 2. The being is in a phase of learning.
> 3. The being receives certain associative stimuli.

Earlier unexpectedness was said to claim focus on the particular incoming content. Based on the above conditions, it's now possible to draw a relation between unexpectedness and learning.

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

In order to learn a certain stimulus is needed. Of course this stimulus may be provided by processes external to the being, or perhaps by the processes that are part of the being itself.

**statement [stimulus]** A stimulus is either provided by the physical world or induced by unexpectedness.

An interesting question is formulated by considering learning. Learning has its complement: forgetting. Forgetting can have many causes, but sometimes a being overrules processes that have become reflexive to the being. These processes may then be replaced by other new processes.

**assumption [embedded processes]** Forgetting old embedded processes means learning new embedded processes.

When a process becomes natural to a being this means that it is ingrained in the beings structure. The process becomes a reflex. A reflex is hard to overrule. Bad habits are said to die slowly, but good habits, if they are truly habits, are just as hard to extinguish.

**statement [embedding of processes]** In cognitive processes slow learning and slow forgetting leads to embedding of certain processes because of rigidity of the learning structures.

When a certain decision is made by the cognitive process this will induce different other processes. Because there are multiple parallel processes that try to induce a different action, not all can carry out the decision they made. For instance when a reflex is too strong, the will cannot succeed. But in many cases beings can be learned to overrule the reflex, and carry out their more contemplated decisions. Both the reflex and the more contemplated process utter a desire to carry out certain decisions. This desire isn't a felt desire, it's just a name to show that the actual decision should, but isn't always carried out. Every decision that's made has to take into account previous decisions.

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

From different regions in our perception we receive different qualities. Different regions we send qualities to may be discerned as well. For different regions certain amounts of competition will take place between different processes. These will all utter their own desires and these will have to enter into competition with each other. Based on this competition the final desire will be derived (which may be a compromise).

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

What the eventual idea is, is that the desire that's uttered tries to force a certain circumstance onto the reference frame corresponding to its value. It can do so by letting the desire stimulate two things. The desire may stimulate perception of certain contents. This by enhancing the values which have actually been perceived. Another thing it may do is stimulate certain processes corresponding to the reference frame. These may then change the reference frame outside of the currently considered stack of processes.

**statement [desire stimulation]** The resulting desire tries to stimulate:
1. the heaviness of content or
2. other processes corresponding to the reference frame.

In order to influence the reference frame it's also important that a being can emit qualities into the reference frame. These will have the same wandering properties as the dormants, and can thus be said to have inherited those properties from dormants. They just add another property to their essence, namely a different origin.

**definition [actuator]** An actuator is a dormant that's emitted by a being.

These actuators can change the reference frame in different ways. Although desires may be used internally to enhance or repress certain incoming contents, the contents can be blocked externally as well. So different processes in the reference frame can be stimulated to free or block receptors. This will then automatically enhance or repress the number of activators.

**statement [freeing and blocking receptors]** In order to enhance or repress the number of activators in a certain time-frame, actuators specialised for respectively freeing and blocking the receptor may be emitted into the reference frame.

In order to be able to reason it's important to have a certain amount of resistance for incoming content. All content should be considered, and based on all content only, can a decision really be made. Although this is a bit of an ideal picture of how reality works, the basic processes do enforce this principle.

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

The above theory links the reference frame and thought together. The theory itself sprung forth from thought. Thought itself sprung forth from the reference frame and its subsequent layers. It's to no avail to start a discussion on whether we should start explaining consciousness from perception itself or the reference frame that is

perceived. Perception cannot be without the reference frame and we can't discuss without perception. They can only be discussed in recurrence: without a beginning and without an end.

> **statement [explaining consciousness]** In explaining consciousness, it doesn't matter if you start at the reference frame or the image we have of the reference frame, since they are both equal in the explanation that should be given of their representations, and thus relational structures.

## Conclusions and recommendations

This article introduces a basic logically minimal model describing the general theory of consciousness. A small extension has been made to the original theory by adding the abstractly defined quality. This should further clarify the contents of the theory, and provide people with an aid to relate different parts of the model to each other.

The original definitions have been rewritten to suit the needs of this new abstract type. Although the original model was also a logically minimal model, this new model with its extension is again a minimal model. None of the definitions can be derived from one or more of the others.

There are still a few things that need to be, or can be, done. First of all a graphical model should be conceived of to describe the theory. Second a logical model has to be constructed in order to prove derived theories. Third a mathematical model should be derived in order to describe the behaviour of the theory. Only when these three models have been thought of can the theory be put to its utmost use in describing the basic principles of consciousness.

## References

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

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

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

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, 2004c.

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

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

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

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