Affective Active Inference and Precision Estimation

A representation of affective feelings in active inference context

S.J. Timmer
4498151

Master of Science Thesis

Delft University of Technology

Delft Center for Cognitive Robotics

Affective Active Inference and Precision Estimation

A representation of affective feelings in active inference context

Master of Science Thesis

For the degree of Master of Science in Mechanical Engineering at Delft University of Technology

S.J. Timmer
4498151

April 17, 2024

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

Abstract

As Neuroscience progresses, there is an increasing amount of research that endorses predictions and reducing of prediction errors as one of the main functions of the brain. active inference is a brain-inspired, mathematical framework that successfully implements this idea both in simulations as well as in robotics. The predictive nature of active inference might make current artificial intelligence agents more adaptive. However, the motives of these agents are often still hardwired as attractor dynamics or learnt using over-engineered rewards. Nature has come up with a different way of providing intelligent beings with drives for their actions: affect, more commonly known as emotions.

Although various models integrate affect into active inference , none have yet applied Mark Solms' definition within a continuous active inference framework. According to Solms' interpretation, affect acts as an evaluative monitoring mechanism of an organism's homeostatic states and guides it through unpredictable environments. This active monitoring of homeostatic states is what according to Solms stands on the basis of consciousness. Key here is the prioritization of different homeostatic needs, where deviations in the most salient category of need come to the organism's affective(conscious) awareness. Mark Solms proposes that computationally, affect is constituted by the inference of changes in precision. Where increases in precision are positively- and decreases in precision are negatively valenced. This change in precision is obtained by performing a gradient descent on free energy with respect to precision, which results in an incremental precision updating scheme that determines the salience of prediction errors. This offers an adaptable mechanism that allows context, through precision modulation, to determine the relative influence of prediction errors.

This in turn allows an agent to prioritize homeostatic needs i.e. letting certain needs come to conscious awareness. "Context" in the light of Solms' research is defined as either: the relation of needs with respect to other needs or the relation of needs with respect to external opportunities. This research supports Solms' theory on affect and consciousness by successfully providing a computational implementation that can, through precision optimization, perform the prioritization of needs directed by "context" as just defined. By doing this successfully, this research shows that the principles used could potentially be useful in continuous active inference implementations, improving their adaptability.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This is document is a part of my Master of Science graduation thesis for the Mechanical Engineering - Biomechanical Design, study programme. The thesis subject was proposed by my daily supervisor C. van Hoof

I would like to thank my supervisors C. van Hoof and prof.dr.ir. M. Wisse for their assistance during the writing of this thesis

Delft, University of Technology
4498151
April 17, 2024

S.J. Timmer

Chapter 1

Introduction

1-1 Introduction

Artificial intelligence is becoming more relevant. While developing these artificial intelligence techniques, neuroscience can be of great value. In neuroscience, there is increasing evidence that prediction is one of the core functions of the brain. Neuroscientist Karl Friston has developed a sophisticated framework that potentially explains the brain's prediction mechanisms [1]. Furthermore, this framework has been applied successfully in simulations [2], as well as in robotics [3]. However, in most of these implementations, motives and goals are pre-determined and fixed using rewards or preferred observations. When looking at ourselves, it is not hard to see that our intelligence or consciousness does not only depend on pre-determined rewards.

One proposition about what drives our behaviour and helps us to maintain our bodily functions by fulfilling its needs, is described in [4] and [5]. In these works, Mark Solms proposes that these behavioural drives are constituted by affect, or more generally known: feelings. Affect is what helps us make good choices, in order to maintain our bodily functions. For example, feeling hungry is an indication that one is deviating from their biological goal. A very important thing to note here is, that affect can even determine behavioural drives, in situations that one has never encountered before. In short, according to Mark Solms, affect is a fundamental signal that evaluates how well we are doing with respect to our biological goals, by measuring how far we are deviating from them and organizing this information in such a way that these deviations will be corrected. Solms even states that affective feelings might be the fundamental source of consciousness [4].

Implementing affective feelings into a model or robot might result in an improvement in their adaptability to novel situations, just like humans. There are already implementations of affect in active inference . However, none of these implements Mark Solms' definition of affect in a continuous active inference model. This research aims to provide a proof of concept for such an implementation. The first part of this research is a deep dive into Mark Solms' research, the second part implements Mark Solms' definition of affect in a continuous active inference model.

1-2 Research Question

As mentioned in the introduction, affect is a mechanism that provides organisms with an adaptative advantage. It might therefore be of interest to implement this in an active inference framework. Mark Solms' proposals [4, 5, 6], provide a basis for the research. According to Solms (see Chapter 3), affect is closely related to changes in free energy, caused by changes in the precision of prediction error.

This research focuses on answering the following research question:

- Can Solms' theory on affect be implemented in a continuous active inference Model?

A series of 5 sub-questions has been established to obtain an answer to this question. The process of answering these 5 questions will provide an answer to the main question:

1. What is affect in both a neurological and computational sense?
2. What is the function of affect in both a neurological and computational sense?
3. What simulation results are needed to prove the model works according to Solms definition?
4. What design specifications does the model need to demonstrate affect successfully?
5. Can this be showcased in a simulation

1-3 Summary of the research

By answering the first sub-question it will be highlighted that affect can be summarised as an evaluative monitoring mechanism of an organism's homeostatic states. Here, deviations from a homeostatically preferred state (increasing prediction error) are evaluated as negative affect. Returning towards its homeostatic preference (decreasing prediction error) is regarded as positive affect. However, the mere presence of affect does not instantly give rise to the feeling of them. Feelings do arise through the inference of changes in expected free energy, or more simply inferences about the uncertainty of the experienced external world and internal body. Computationally, this inference is done by precision estimation. Here decreasing precision (increasing uncertainty) is registered as unpleasure and increasing precision (decreasing uncertainty) is registered as pleasure. It will be shown that the inferences of precisions are done by performing a gradient descent of free energy with respect to precision. The inference of this precision in interoceptive context is what gives rise to felt uncertainty, or 'affective consciousness. The same mechanism in the exteroceptive world is what gives rise to 'perceptual consciousness' or attention.

In Section 3-3, it will follow that precision (and thus affect) plays a crucial role in the prioritization of homeostatic needs, by acting as a gain modulator on prediction errors. This addresses the second sub-question. Prioritization of need adds adaptive value to the organism as it gives the ability to take context into account. This prioritization according to context happens in two ways:

1. Precision (and thus affect) contextualises and prioritises needs in relation to other needs.
2. Precision contextualises needs in relation to external opportunities.

For a model to comply with Solms definition of affect, it needs to include an implementation of precision optimization. On top of this it needs to be able to perform need prioritization in the two ways as described above. This addresses the third sub-question and is discussed in Section 3-4. Chapter 4 will then look more into the specific model design requirements for it to successfully perform the two prioritization tasks. A simulation for a microscopic animal called 'Hydar' is constructed. The goal of Hydar is to use affective active inference to prioritize either his need for food or his preferred temperature by modulating the precisions of signals. For this to work, the precision optimization formula, as described in Chapter 3, needs to be implemented. This is an important requirement as this is a key aspect of affect.

On top of this, there are 4 concrete characteristics in the model architecture that need to be present:

1. Hydar needs **multiple** interoceptive needs that can compete.
2. Hydar needs an **interoceptive** system that can measure and prioritize interoceptive prediction errors.
3. Hydar needs **exteroception** to prioritize these needs according to context.
4. Hydar needs **action**

Together these model characteristics provide an insight into the fourth sub-question. Chapter 5-2 provides two simulations that illustrate Hydar performing the two prioritization tasks using precision optimization addressing the final sub-questions.

Altogether, this research storyline provides support for Mark Solms' description of affect in an active inference context. This is namely because the results show that using the computational definition of affect as described by Solms, it can successfully perform the prioritization tasks that are, according to Solms, the reason affect exists in the first place.

This then relates to the main research question. The research shows that indeed that affect as described by Solms can be successfully implemented in a continuous active inference model. This provides support to Solms' definition of affect. The implementation of this research serves as a proof of concept, rather than a full affective active inference model. That is, the model is of simple structure and behaves in a heavily simplified environment. Nonetheless, it provides promising results and sets up the potential for further research, using possibly extended models. Examples of this would be the implementation of a more extensive hierarchy or more complex order relations in the model dynamics. This is discussed in Chapter 6.

Chapter 2

Active inference

2-1 What is active inference ?

Before starting the discussion about how affect can be implemented in active inference , it needs to be clarified what active inference is and how it works. This chapter will mostly be devoted to explaining active inference in a more mathematical context. Firstly, a short, high-level, general description.

2-1-1 Introduction to active inference

Active inference is about the minimization of a quantity called free energy. Free energy is the information-theoretic analogue of entropy in thermodynamics. Free energy describes the amount of disorder in the internal and external states of an organism (or an artificial intelligence agent). States of an organism are sensed by the sensory system. Sensory states are comprised of all the different ways that an organism can infer the "hidden states of the world". For example, eyesight is a representation of the real world in your mind, based on information caused by photons hitting your retina. In other words, what you see is not the "real world", it is merely the representation that you can perceive of a hidden state (note that an organism's internal states can also be hidden from the brain, e.g body temperature can only be inferred through a signal provided by thermoreceptors in one's body). The organism also has an internal "generative model", which is based on earlier experiences. This generative model can make predictions about the states that the organism wants to infer. This prediction is continuously compared to the sensory information in order to make an as good as possible estimation about the "real world's" hidden states. The difference between the generative models' prediction and sensory states is called the prediction error. The better the model's representation of the real world, the better the predictions and thus the lower the free energy. In order to survive, the organism wants to minimize free energy, by minimizing this prediction error. It does so in different ways. The first way of minimizing free energy is by optimizing predictions such that they match sensory input. The second way of minimizing free energy is through action. In other words: manipulating the environment in such a way that it matches expectation.

In its full form, active inference is not limited to just one level that is minimizing free energy. Instead, it is set up in a hierarchical way, inspired by the brain [7]. The idea behind this hierarchical structure is that higher layers, dealing with more abstract information send down prior predictions to lower levels that are concerned with more concrete sensory information. The lower levels use these priors to compute prediction errors, which are then sent up the hierarchy again. This prediction error is then used as information to update the generative model of the layer above. This hierarchical layering is inspired by the way neurons in the human cortex are wired.

Active inference frameworks can be represented in the continuous domain, as well as the discrete domain. While they are both centered around the minimization of free energy, they do encompass two different worlds. Continuous active inference has now been introduced. This is what will be used in this research. The following sections provide a more in-depth view into Continuous active inference . For more details about the mathematics involved one can consult [8].

2-2 active inference

2-2-1 Free Energy

Active inference revolves around the minimization of free energy. Free energy can be seen as a bound on surprise, which is inherent in sensory data [1]. The free energy equation can be derived using the Kullback-Leibler (KL) divergence which results in the following definition [7]:

$$F = -\ln p(y) + D_{KL}(q(\vartheta;\lambda) \| p(\vartheta \mid y)) \tag{2-1}$$

Here $p(\vartheta \mid y)$ corresponds to the posterior that the agent needs to approximate with $q(\vartheta;\lambda)$. Furthermore, $p(y)$ denotes the surprise. The equation can be rewritten into:

$$F = -\int q(\vartheta;\lambda) \ln \frac{p(y,\vartheta)}{q(\vartheta;\lambda)} \mathrm{d}\vartheta \tag{2-2}$$

The key concept of the free energy formula is that it contains an arbitrary distribution, with a chosen mean and variance, such that it approximates the intractable posterior $p(\vartheta|y)$. Here, ϑ denotes the hidden states that the agent wants to estimate and y denotes the sensory evidence that the agent has access to, in order to do so. This approximation for $p(\vartheta|y)$, is denoted in equation 2-1 as $q(\vartheta;\lambda)$ and is called the recognition density. It is used to estimate the system's hidden states, which are described by ϑ, parameterized by the arbitrarily chosen mean and variance, denoted as λ. As can be seen in the second term of Equation 2-1, these two distributions are being used in a KL-divergence, which quantifies the difference between them. The first term in Equation 2-1, denotes the surprise over the sensory input [7]. This part can be minimized through action u.

So, in short, the KL-divergence is used to estimate a distribution that can approximate the posterior and thus can estimate the hidden states, given sensory input $\tilde{y}$. However as mentioned above, this posterior is often intractable and thus, to use this principle the free energy equation needs to be rewritten such that the arbitrary estimate can be 'fitted' to a distribution that can be quantified. To do so, Equation 2-1 can be rewritten into Equation 2-2. This equation consists of the aforementioned recognition density and also $p(\tilde{y}, \vartheta)$, which is called the generative density. This generative density, or generative model is the brains representation of the real world and is something that can be quantified. Using the product rule, it can be divided into two terms:

$$p(\tilde{y}, \vartheta) = p(y|\vartheta) * p(\vartheta) \tag{2-3}$$

Here,$p(\vartheta)$, represents the state dynamics over time and $p(\tilde{y}, \vartheta)$ provides a probabilistic mapping from the states to the sensory input. This will later be explained in more detail.

What is for now important, is that the free energy Equation 2-2, measures the difference between the recognition density and the brain's generative model. The main goal of active inference is the minimization of this free energy. This can be done in two different ways:

1. Improving the system's internal model and perception
2. Acting upon the environment in such a way that it lowers surprise.

Knowing ow what all the elements of Equation 2-2 and 2-1 represent, it can be simplified under the Laplace assumption. The full mathematical description can be found in [8], but goes beyond the purpose of this text. The crux of this assumption is however that Gaussian densities that are sharply peaked around its mean: μ. Thus the the estimation of the recognition density is reduced to finding the proper mean μ.

The recognition density and free energy respectively can be simplified to:

$$q(\vartheta; \lambda) \approx q(\vartheta; \mu) = \mu \tag{2-4}$$

$$F \approx F(y, \mu) = -\ln p(y, \mu) \tag{2-5}$$

One thing that that needs to be noted now, is that by using the Laplace assumption, variance is not gotten rid of altogether. Under this assumption, estimating the mean μ of the hidden states θ will provide enough information to accurately compute free energy. There will however still be variance when this mean is compared to prior expectations, by calculating prediction error. This will become more clear in the next paragraphs

For now, it suffices to say that free energy minimization is now reduced to iteratively finding a μ, that produces the lowest free energy, where the exact metric that is used for this scheme is the prediction error. This will be explained in the next subsection.

2-2-2 Prediction Error Minimization

Equation 2-5 can be split up into two components:

$$F(y,\mu) = -\ln p(y|\mu) - \ln p(\mu) \tag{2-6}$$

Equation 2-6, shows that there is a density for sensory mapping and a density for the state dynamics just as described in Equation 2-3. Another way to describe these generative models is by using a state-space representation.

$$\begin{aligned} \mu &= \overline{\mu} + w \\ y &= g(\mu) + z \end{aligned} \tag{2-7}$$

Here, the first equation describes how the agent believes that hidden states, as described by μ, are generated. Here $\overline{\mu}$ is a prior estimation of the hidden states and w is random noise with zero mean and variance σ_w^2. So in other words, the agent believes that the mean μ of the actual hidden states θ fluctuates around its prior belief $\overline{\mu}$. The second equation describes how the agent's internal model prediction of the sensory input using $g(\mu)$ compares to the actual sensory input y, where again the sensory states fluctuate around an estimate that relates y with μ. This fluctuation is again described with a noise term z with zero mean and variance σ_z^2.

Now, coming back to the earlier note on variance and the Laplace Assumption, note needs to be taken that σ_w and σ_z differ from the variance that is neglected under the Laplace Assumption. Variance under Laplace Assumption refers to the arbitrary statistics of the recognition density, whereas σ_w and σ_z refer to the variability of estimations of μ compared to the prior on the mean $\overline{\mu}$ and the variability between sensory states y and predicted sensory states by the internal model $g(\mu)$. As the noise terms are assumed to be Gaussian, the equations can be written in the following way:

$$\begin{aligned} p(\mu) &= \frac{1}{\sqrt{2\pi\sigma_w^2}} e^{\left\{-(\mu-\bar{\mu})^2/\left(2\sigma_w^2\right)\right\}} \\ p(y \mid \mu) &= \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{\left\{-(y-g(\mu))^2/\left(2\sigma_z^2\right)\right\}} \end{aligned} \tag{2-8}$$

Filling these two results into Equation 2-6 obtains:

$$F(y,\mu) = \frac{1}{2}\left(\frac{1}{\sigma_z^2}\varepsilon_y^2 + \ln\left(\sigma_z^2\right)\right) + \frac{1}{2}\left(\frac{1}{\sigma_w^2}\varepsilon_x^2 + \ln\left(\sigma_w^2\right)\right) \tag{2-9}$$

with

$$\begin{aligned} \mu &= \overline{\mu} + w \\ y &= g(\mu) + z \end{aligned} \tag{2-10}$$

Thus the prediction errors are defined as:

$$\begin{aligned} \varepsilon_x &= \mu - \overline{\mu} \\ \varepsilon_y &= y - g(\mu) \end{aligned} \tag{2-11}$$

Now, looking at Equations 2-9 and 2-11, it is evident that the minimization of free energy comes down to the minimization of two prediction errors: ε_y error measures the difference between the sensory data and the part of the generative model that accounts for sensory mapping. The other error ε_x measures the difference between some sort of "belief of the hidden states" and the function of motion that is incorporated in the brain's generative model.

In order to be able to describe this in a computational model, the concept: "belief of the hidden states" or "prior" needs some further explanation, which will be done in the following subsection.

2-2-3 Hierarchical message passing

As of today, the cortex is regarded as a hierarchical system [7]. It consists of billions of neurons and their interconnections add up into trillions. One key aspect of active inference is how these connections are structured. The cortex consists of multiple specialized areas, that can be subdivided into new areas up until one reaches the so-called "Micro-Columns". Cognitive processes are the product of these areas working together. Each area is specialized and has to deal with its own type of information. As a result, some areas will be closer to concrete sensory information, whereas others will be more associated with more abstract information. According to many researchers [7], it is believed that bottom-up and top-down connections are structured in such a way, that sensory information obtained in lower levels can be sent upwards in the form of prediction errors, whereas more abstract information generated in the higher levels can be sent downwards in the form of prior beliefs.

In [8] a simple hierarchical scheme would look like this:

$$\begin{aligned} y &= g^{(1)}\left(\mu^{(1)}\right) + z^{(0)} \\ \mu^{(1)} &= g^{(2)}\left(\mu^{(2)}\right) + z^{(1)} \\ \mu^{(2)} &= \cdots \\ &\vdots \\ \mu^{(M)} &= z^{(M)} \end{aligned} \tag{2-12}$$

or more compactly as:

$$\mu^{(i-1)} = g^{(i)}\left(\mu^{(i)}\right) + z^{(i)} \tag{2-13}$$

The key point about the hierarchical structure is that a state estimation μ on a layer will be used as "sensory" information in the layer above. The alert reader would now see that this structure is simply Equation 2-7 with more layers on top of each other. In other words, it can be seen as a split version of Equation 2-9 that is stacked on top of each other (without the sigma terms). It is important for the reader to keep this hierarchical structure in the back of his head as it will be used in the final model.

When incorporating this form of hierarchical scheme in a dynamic model, it can be expanded into:

$$\begin{aligned} D\tilde{\mu}_x^{(i)} &= \tilde{f}\left(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}\right) + \tilde{w}^{(i)} \\ \tilde{\mu}_v^{(i-1)} &= \tilde{g}\left(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}\right) + \tilde{z}^{(i)} \\ \tilde{y} &= \tilde{\mu}_v^{(0)} \end{aligned} \tag{2-14}$$

This hierarchical structure will also be used in the final model and thus needs some more elaboration:

As can be seen, there are multiple things added to the original state space 2-7. First of all, all the variables contain a"˜" operator which is short for a vector containing derivatives until order n. With this also comes the D operator, which indicates that all the derivatives in the vector are moved one place upwards. This results in a form $D\tilde{\mu}_x = \tilde{f}(\tilde{\mu}_x) + \tilde{w}$, where the derivatives of μ_x are coupled with some function: $\tilde{f}(\tilde{\mu}_x^{(i)})$. This creates a dynamic coupling over time.

The two major differences however are the introduction of $\tilde{\mu}_v$ and the (i) index. The index represents the specific layer that the generative model belongs to. Each layer provides a prior $\tilde{\mu}_v$, to the layer that is one step down in the hierarchy. This prior can then be used in the generative model of that next layer and so on. Likewise, sensory information can be sent upwards in these layers in the form of prediction errors. Using this scheme, Equation 2-11 can be expanded creating the following prediction error equations:

$$\begin{aligned} \tilde{\varepsilon}_x^{(i)} &= D\tilde{\mu}_x^{(i)} - f(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}) \\ \tilde{\varepsilon}_v^{(i)} &= \tilde{\mu}_v^{(i-1)} - g(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}) \end{aligned} \tag{2-15}$$

Figure 2-1, shows a scheme that illustrates these top-down and bottom-up pathways for the priors and prediction errors.

Using the new hierarchical structure, the free energy equation can be rewritten as follows:

$$F^{(i)}(\tilde{y}, \mu) = \frac{1}{2}(\tilde{\varepsilon}_x^{(i)T}\tilde{\prod}_w^{(i)}\tilde{\varepsilon}_x^{(i)} + \tilde{\varepsilon}_v^{(i)T}\tilde{\prod}_z^{(i)}\tilde{\varepsilon}_v^{(i)} - \ln|\tilde{\prod}_w^{(i)}| - \ln|\tilde{\prod}_z^{(i)}|) \tag{2-16}$$

Where $\tilde{\prod}$, which is called the precision matrix, denotes the inverse of the covariance matrix. For exact derivation, [8] can be consulted.

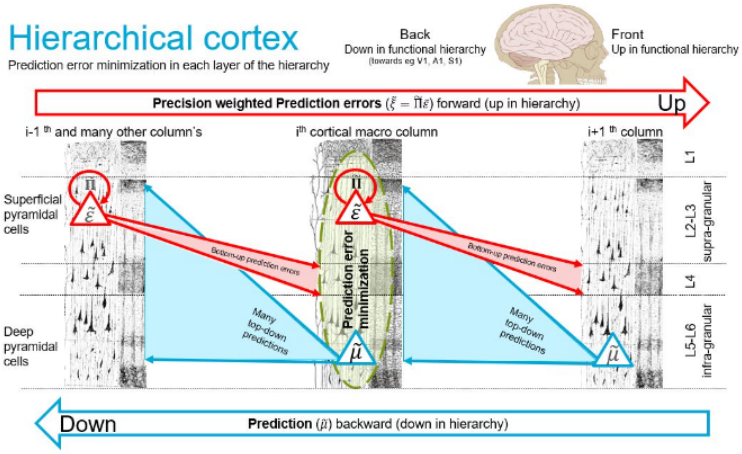


Figure 2-1: A visual representation of the hierarchical structure in active inference . Predictions are sent down the hierarchy and prediction errors are sent up the hierarchy. Source: https://www.kaggle.com/code/charel/learn-by-example-active-inference-in-the-brain-3

2-2-4 Gradient Descent

As mentioned earlier, active inference revolves around the minimization of free energy. The first two variables that can be optimized over are the beliefs of the hidden states and the priors:

$$\begin{aligned} \tilde{\mu}_x^{(i)} &= \underset{\tilde{\mu}_x^{(i)}}{Argmin}\, F(\tilde{y}, \mu) \\ \tilde{\mu}_v^{(i)} &= \underset{\tilde{\mu}_v^{(i)}}{Argmin}\, F(\tilde{y}, \mu) \end{aligned} \tag{2-17}$$

This process is done through the process of gradient descent. Gradient descent allows the agent to optimize over the different variables that make up the free energy equation.

$$\begin{aligned} \dot{\tilde{\mu}}_x^{(i)} &= \tilde{\mu}_x^{\prime(i)} - \frac{\partial F(\tilde{y}, \mu)}{\partial \tilde{\mu}_x^{(i)}} = D\tilde{\mu}_x^{(i)} - \frac{\partial F(\tilde{y}, \mu)}{\partial \tilde{\mu}_x^{(i)}} \\ \dot{\tilde{\mu}}_v^{(i)} &= \tilde{\mu}_v^{\prime(i)} - \frac{\partial F(\tilde{y}, \mu)}{\partial \tilde{\mu}_v^{(i)}} = D\tilde{\mu}_v^{(i)} - \frac{\partial F(\tilde{y}, \mu)}{\partial \tilde{\mu}_v^{(i)}} \end{aligned} \tag{2-18}$$

When filling in Equation 2-16 and applying some differentiation the following two equations are obtained:

$$\dot{\tilde{\mu}}_x^{(i)} = D\tilde{\mu}_x^{(i)} - \frac{\partial \tilde{\varepsilon}^{(i)\top}}{\partial \tilde{\mu}_x^{(i)}} \tilde{\xi}^{(i)} \tag{2-19a}$$

$$\dot{\tilde{\mu}}_v^{(i)} = D\tilde{\mu}_v^{(i)} - \frac{\partial \tilde{\varepsilon}^{(i)\top}}{\partial \tilde{\mu}_v^{(i)}} \tilde{\xi}^{(i)} - \tilde{\xi}_v^{(i+1)} \tag{2-19b}$$

Where the precision weighted prediction error ($\tilde{\xi}^{(i)} = \tilde{\Pi}^{(i)}\tilde{\varepsilon}^{(i)}$), or the so called precision weighted prediction error, this term is relevant as it is used in the bottom-up pathway of the prediction error (see Figure 2-1). These two gradient descent functions are used in active inference to update the belief of the hidden states as well as the prior, such that the free energy is minimized.

2-2-5 active inference

Aside from updating the hidden representations, active inference also involves action u, in order to minimize free energy.

$$u = \underset{u}{Argmin}\, F(\tilde{y}, \mu) \tag{2-20}$$

This, again, is done by gradient descent, this time with a derivative with respect to u:

$$\dot{u} = -\frac{\partial F(\tilde{y}, \mu)}{\partial u} = -\frac{\partial \tilde{y}}{\partial u}^{\top}\frac{\partial F(\tilde{y}, \mu)}{\partial \tilde{y}} \tag{2-21}$$

Important to note here, is that the free energy Equation 2-16 does not depend directly on u, which is why Equation 2-21 has to be split in the two partial derivatives. The last of these can be derived into:

$$\frac{\partial F(\tilde{y}, \mu)}{\partial \tilde{y}} = \frac{\partial F(\tilde{y}, \mu)^{(1)}}{\partial \tilde{y}} = \frac{\partial \tilde{\varepsilon}_v^{(1)}}{\partial \tilde{y}}^{\top} \tilde{\Pi}_z^{(1)}\tilde{\varepsilon}_v^{(1)} = \tilde{\Pi}_z^{(1)}\tilde{\varepsilon}_v^{(1)} = \tilde{\xi}_v^{(1)} \tag{2-22}$$

$\dot{u}$ can then be written as:

$$\dot{u} = -\frac{\partial \tilde{y}}{\partial u}^{\top}\tilde{\xi}_v^{(1)} \tag{2-23}$$

Analogously to the gradient descent in sub section 2-2-4, Equation 2-21, optimizes the action variable u, so that free energy is minimized. The real life representation of this process, would be that the organism uses action to manipulate the environment in such a way, such that uncertainty minimized. For example, an organism that expects to maintain a certain body temperature by moving from a cold to a warmer place in order to keep that temperature.

Chapter 3

Affect as defined by Solms

3-1 Introduction

With active inference and the free energy principle covered, the next step is defining affect. The entirety of this chapter is based on the interpretation and function of affect that is proposed by Mark Solms. In his scientific collaborations with Friston [4, 5] and the book [6], Solms describes what affect is, how it is linked to consciousness and how it fits in the active inference context.

There are already various implementations of affect in active inference , such as [9, 10, 11], but these do not focus on Solms' definition of affect in a continuous active inference domain. The goal of this research, as clearly described by the main research question, is to provide a computational scheme that successfully implements an affective mechanism as described by Solms in [4, 5, 6] in a continuous active inference framework. This computational scheme can then be used to provide simulation results, which can support Solms' theory on affect. Furthermore, it can provide a basis for using precision optimization for the benefit of adaptability in an active inference framework. Before this is done, it needs to be clear how affect is defined both neurologically as well as computationally, what its function is and lastly what a simulation needs to be able to do to work conform to Mark Solms' definition of affect. Doing this, this chapter focuses on answering the first three research sub-questions:

1. What is affect in both a neurological and computational sense?
2. What is the function of affect?
3. What simulation results are needed to prove the model works according to Solms definition?

The first two sub-questions highlight two crucial aspects of Mark Solms' definition of affect. The third question is aimed at synthesising these aspects to form a complete definition of affect. This will be regarded as the requirement for the simulation:

1. Affect comes forth from the inference of uncertainty about an organism's ability to comply with its own homeostatic needs. This inference determines the valence of the affective feelings that an organism experiences. Here deviations from a homeostatic equilibrium (increasing uncertainty) is registered as negative affect. Moving towards a homeostatic equilibrium (decreasing uncertainty) is registered as positive affect. Computationally uncertainty is inferred by precision estimation, which is done by performing a gradient descent of free energy with respect to precision.

2. The function of affect is the prioritization of different homeostatic needs. This is where the gain modulation property of precision in active inference comes into play, as it influences the magnitude of the error signal. In other words, it influences 'arousal'. Here deviations from a desired state in the most salient category of need, that is, the category with the highest precision come to affective awareness. In other words these become conscious. The prioritization of each category is determined by context. This contextualization can take two forms:

   (A) The prioritization of needs in relation to other needs.

   (B) The prioritization of needs in relation to external opportunities and restrictions.

   Precision plays a crucial role in both these mechanisms. First of all, needs with a high afforded precision are prioritized over precisions with low afforded precision. Secondly, precision plays a role as exteroceptive information with higher precision will be prioritized over exteroceptive information with lower precision.

These two key aspects of affect are summarized in Figure 3-1. In [5], Pfaff's analogy of affect with a vector is mentioned. Here valence is analogous to the angle of the vector. Arousal is analogous to the magnitude of the vector.

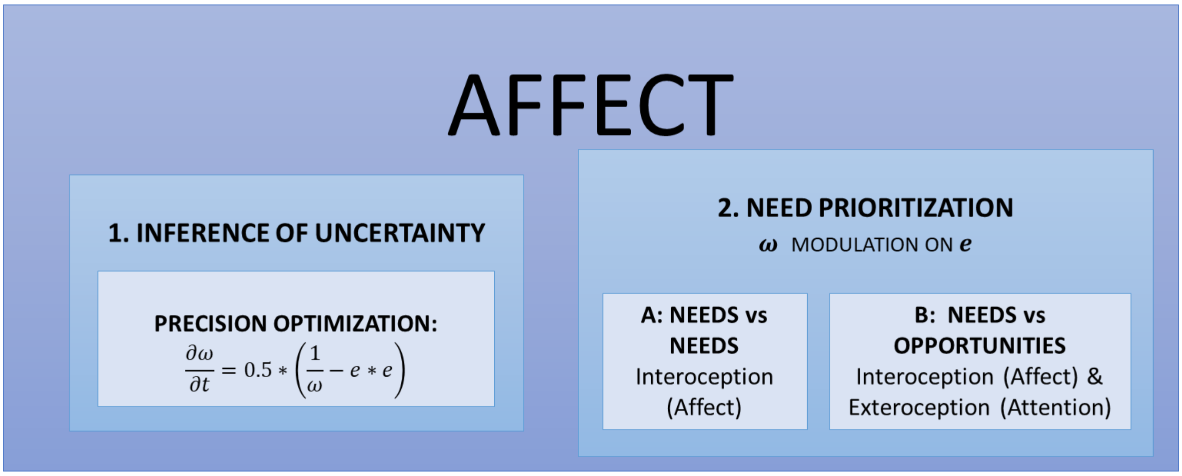


Figure 3-1: Affect as described by Solms and has two crucial components. 1. Precision optimization 2. Need prioritization. The second aspect takes form in two ways: A: The prioritization of needs in relation to other needs. B: The prioritization of needs in relation to external opportunities (and restrictions).

This chapter dives deeper into these two aspects. Each section covers one aspect and provides an answer to one of the first two sub-questions. The following section focuses on the synthesis.

After this, it should be clear which requirements are needed to successfully demonstrate affect in an active inference simulation.

3-2 Affect and precision optimization

3-2-1 Section overview

Addressing the first research sub-question, it will become clear in this chapter that affect can be summarised as an evaluative monitoring mechanism of an organism's homeostatic states. Here, deviations from a homeostatically preferred state (increasing prediction error) are evaluated as negative affect. Returning towards its homeostatic preference (decreasing prediction error) is regarded as positive affect. These deviations do not instantly give rise to the feeling of affect. Feelings do arise through the inference of changes in free energy, or more precisely, inferences about the uncertainty of the experienced external world and internal body. Computationally, this inference is done by precision estimation. Here decreasing precision (increasing uncertainty) is registered as unpleasure and increasing precision (decreasing uncertainty) is registered as pleasure. In other words, the inference of changes in precision determines whether an organsism experiences positively or negatively valenced affect (Figure 3-2). This section will show that the inferences of precisions are done by performing a gradient descent of free energy with respect to precision and will therefore be mainly focused on the first part of Figure 3-2, which associated with valence (the angle of the vector [12]). The inference of this precision in interoceptive context is what gives rise to "affective consciousness". The same mechanism in the exteroceptive world is what gives rise to "perceptual consciousness" or attention.

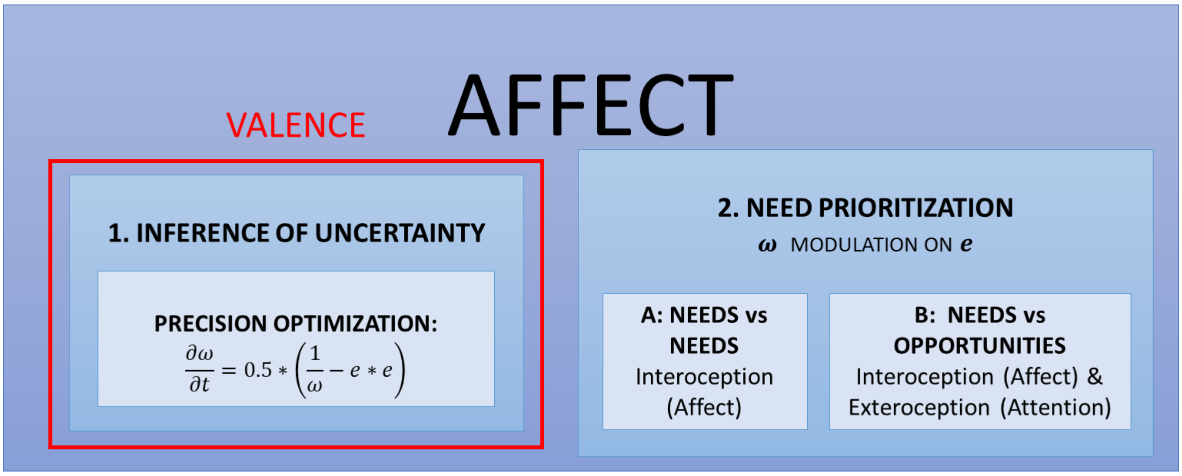


Figure 3-2: The focus of this section will be on the first aspect of affect: the inference of uncertainty through precision optimization. This part of affect is associated with valence (whether the affect is positive or negative)

The section afterwards will focus more on the second part of the story is that this inferred precision dictates the extent to which different prediction errors are weighted among one another. In other words, to what extent certain prediction errors, which represent (uncertainty in) an organism's homeostatic needs [6], are felt. In short this can be defined as arousal (analogous to the magnitude of the vector ([12]). This weighting of prediction errors fulfils the function of prioritization of internal needs and external opportunities, which will be also be elaborated on in Section 3-3.

3-2-2 How do we feel?

Introduction

The works of Solms that are considered for this thesis are all centered around the "Hard problem of consciousness" as formulated by Chalmers [13]. The problem can be described in one sentence like this: If experience (consciousness) arises from physical processes (i.e. electrical signals in neurons), how and why does it?

In the texts [4] and [6], Solms describes that various experiments showed that when the brainstem is ablated from the central nervous system, consciousness is completely lost. Conversely, other experiments have shown that when damage is done to parts of the cortex this is not so much the case. Even when the subject shows signs of cognitive decline, a sense of consciousness is maintained. This leads us to believe that, as the brainstem is the organ that is the source of our affective feelings, consciousness is an affective process. Or more explicitly formulated: the arousal process that produces what we call consciousness, is constituted by affect or feeling . Furthermore, [5] states that "the arousal of consciousness and homeostatic regulations are effected by the same parts of the brain".

This relationship between affect and consciousness stands central in Solms' research:

> *"If core brainstem consciousness is the primary type, then consciousness is fundamentally affective. The arousal processes that produce what is conventionally called "wakefulness" constitute the experiencing subject. In other words, the experiencing subject is constituted by affect."* [4]

As for the practical implementation of this extension of Solms' research, affect, and thereby consciousness have yet to be concretely defined. At the basis, all 3 reviewed works pivot around the concept of affect coming forth from the maintenance of homeostasis.

Homeostasis and affect

As described in Chapter 2 the ultimate goal of active inference is for an organism to keep existing. To do so, the organism needs to work against the world's natural tendency towards disorder or entropy. This tendency to disorder is commonly known as "the second rule of thermodynamics". The biological mechanism for resisting this rule is by maintaining homeostasis.

Organism's receive information about their likely survival by asking questions (i.e. taking measurements) of their biological state in relation to unfolding events. The more uncertain the answers are (i.e. the higher the entropy) the worse for the organism; it means it is failing in its homeostatic obligation to occupy limited states (its expected states)[6].

By putting effort into trying to remain at its homeostatic values, the organism resists entropy as it reduces the probability of being dispersed over multiple states which the natural law of entropy tries to dictate. The other way around, deviating from a homeostatic preference indicates a dispersion of possible states and thus an adherence to the natural entropic forces. An organism needs to resist these forces to survive. An alert reader should now recognize that this is a general summary of active inference as specified earlier.

Using this idea, it can be argued that from an existential point of view, deviating away from homeostasis (increasing uncertainty), is objectively "bad" and moving towards its homeostatic preferences (increasing uncertainty) is objectively "good". In his works [4, 6, 5] Solms bases his definition of affects, or feelings on this rule. Here deviations away from homeostasis are registered as negative affect and moving towards it is registered as positive affect. In other words, according to Solms, affect comes forth from the active inference principle and its properties of resisting natural disorder by minimizing free energy. It is the means whereby organisms register their own states [5].

The big question of the Hard Problem [13] however still remains as this explanation does not cover why these deviations and movements towards homeostatic preferences give rise to conscious feelings in the way we experience them. Coming from the arguments just explained, these homeostatic mechanisms could be exerted without any form of consciousness ([4] and [5] refer to this as "philosophical zombies") and still work. Even more so, innumerable homeostatic regulations happen within our bodies, without us noticing. Yet, we (and most likely other organisms) do experience what we call consciousness. Thus the question remains as to how consciousness arises. How is it exactly that we consciously experience these (quantitative) homeostatic deviations as (qualitative) feelings? The answer that Solms proposes to this question is: "through inferring the changes in uncertainty about the homeostatic predictions".

3-2-3 Consciousness and affect

The inference of uncertainty

Incoming homeostatic prediction errors are indications of uncertainty around the state of affairs. Inferences about the fluctuations in these errors, thus provide information for the organism as to whether it is doing well with respect to that homeostatic modality. On top of this, in different situations, different modalities are of importance. That is, in some situations the organism can afford a higher uncertainty on a homeostatic need than in other situations. Therefore the organism needs a way to not only infer changes in uncertainty but also to prioritize which needs it cannot afford to risk increasing uncertainty.

Precision optimization

The formal implementation that addresses these issues in an active inference context is formulated by Solms as precision optimization. As already touched upon in Chapter 2, the precisions are the inverses of variance and represent the agent's confidence afforded probabilistic beliefs about states of the outside world [4]. They play a crucial role in the free energy minimization scheme as they act as a gain modulator on the prediction errors. Precision optimization (or precision estimation) fits well in this free energy minimization scheme as will be elaborated on below.

Homeostasis as described above arises through active inference and free energy minimization as covered in Chapter 2. Here, emphasis lies on two ways by which this could be done. Firstly, an organism can act upon its environment to change its sensations so they match predictions. Secondly, one can change internal representations, to produce a better prediction. As mentioned earlier in Chapter 2, these two quantities that are to be minimized, are modulated with the use of their respective precisions.

The crux of Solms' work lies in the proposition of a third way of minimizing free energy. This third way has been formulated in [5] as: *"Adjusting the precision to optimally match the amplitude of prediction errors"*. This can be done by performing a gradient descent of free energy with respect to precision. This gradient descent, along with the gradient descents on action and Perception as defined in Chapter 2 are formulated in [4, 6, 5] as below in Figure 3-3 and Equations 3-1:

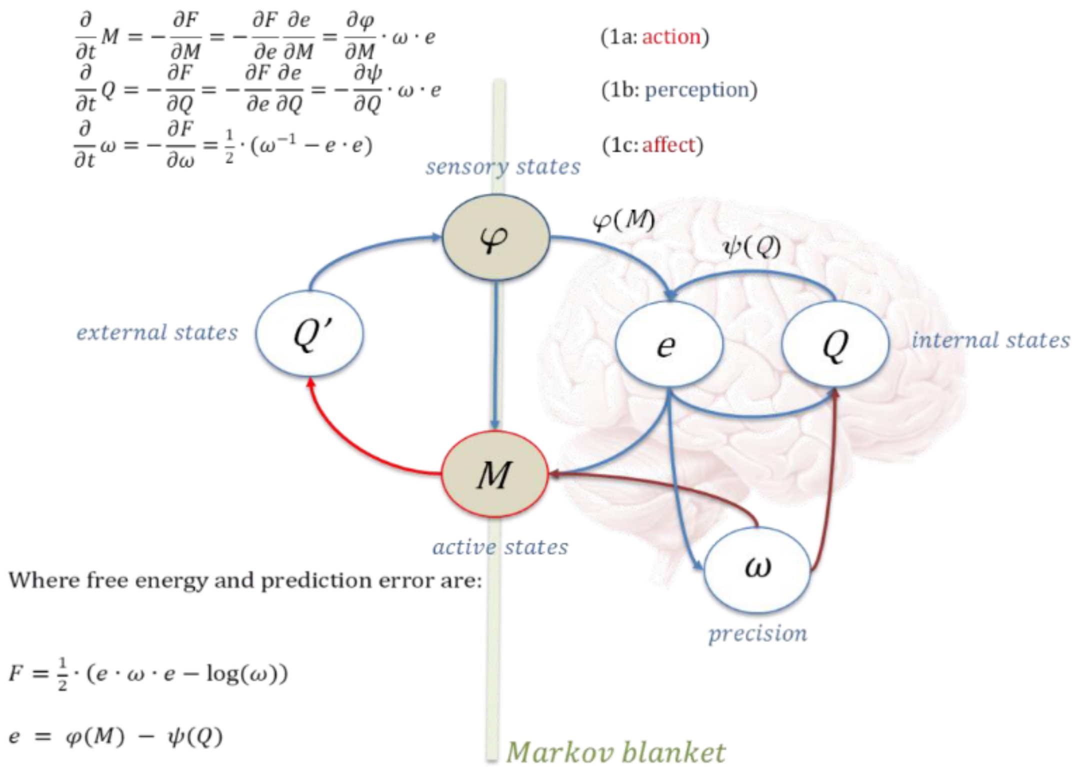


Figure 3-3: In this image, one can see the three gradient descent equations alongside a diagram of a self-organising system using active inference . Crucial is the third equation that highlights the optimization of precision ω which is the process that is associated with affect and consciousness in Solms' works.

The notations of the works [4, 5, 6] have been used here (which are originally derived from notations by Freud [14]). These notations differ from the notations in Chapter 2. For comparison one can consult Table 3-1 and Figure 3-3.

Figure 3-3, shows a schematic illustration of an active inference agent. Here the sensory information φ, informed by an external hidden state Q' which is influenced by action M, gets compared to a generative model's $\psi(Q)$ representation. Here Q denotes the estimate of the external hidden state Q' and $\psi(Q)$ is the internal models ψ expectation of the sensory signal that Q would generate. This comparison between these two variables, results in a prediction error e, which is used to update the internal representation Q, the active states M (which influences the external or vegetative internal world Q') and most importantly for this research in the updating of precision ω.

The gradient descent functions that are used to perform these three updates are listed in Equation 3-1. Here, Equation 3-1a describes the gradient descent function of free energy F with respect to actions M, where φ denotes the sensory states. Equation 3-1b describes the gradient descent of free energy with respect to the internal expectations Q about the external states and ψ is a prediction of the sensory inputs that would have been encountered if the external (hidden) states would have been equal to Q. Equation 3-1 denotes a gradient descent of F with respect to precision ω. It is this equation that Solms associates with affect and consciousness [5].

$$\frac{\partial}{\partial t}M = -\frac{\partial F}{\partial M} = -\frac{\partial F}{\partial e}\frac{\partial e}{\partial M} = \frac{\partial \varphi}{\partial M} \cdot \omega \cdot e \tag{3-1a}$$

$$\frac{\partial}{\partial t}Q = -\frac{\partial F}{\partial Q} = -\frac{\partial F}{\partial e}\frac{\partial e}{\partial Q} = -\frac{\partial \psi}{\partial Q} \cdot \omega \cdot e \tag{3-1b}$$

$$\frac{\partial}{\partial t}\omega = -\frac{\partial F}{\partial \omega} = \frac{1}{2} \cdot \left(\omega^{-1} - e \cdot e\right) \tag{3-1c}$$

Free Energy and prediction error can be described as:

$$F = \frac{1}{2} \cdot (e \cdot \omega \cdot e - \log(\omega)) \tag{3-2a}$$

$$e = \varphi(M) - \psi(Q) \tag{3-2b}$$

The definition of free energy mentioned here is a simplified version of-, but could theoretically be expanded to Equation 2-9 or 2-16. The prediction error as depicted here is comparable to Equation 2-11 or 2-15. As already explained, Equation 3-2 depicts e as the difference between the sensory states φ as a function of action M (where action influences external states and external states influence the sensory states, see Figure 3-3) and the expectation of a sensory signal using $\psi(Q)$.

When looking at Equation 3-1c, one can see that it is dependent on the difference between the inverse of precision ω^{-1} (which is the same as average uncertainty or variance σ^2) and the

Variable meaning	Variable Solms	Variable Chapter 2
action	M	u
Prediction error	e	ε_x & ε_y
Sensory input	φ	y
Precision	ω	$1/\sigma_z^2$ & $1/\sigma_w^2$
Hidden state estimate	Q	μ
Generative model	ψ	$f(\mu)$ & $g(\mu)$

Table 3-1: Variables used by Solms vs. variables in Chapter 2

square of the prediction error e^2. When taking into account that variance can be described as the average of all prediction errors we can rewrite Equation 3-1c into:

$$\frac{\partial}{\partial t}\omega = -\frac{\partial F}{\partial \omega} = \frac{1}{2} \cdot \left(\frac{\sum(\mu - \overline{\mu})^2}{N} - e^2 \right) \tag{3-3}$$

Here we can see that the change in precision is dependent on the difference between variance ($\frac{1}{\omega}$), which is average prediction error, and the prediction error at a specific instance of measurement. When these are equal, we get $\frac{\partial}{\partial t}\omega = 0$. In other words, when the measured prediction error is equal to its average, the average does not need to be changed. However, when there is a difference between the prediction error and its average, the agent needs to account for this by updating the variance through $\frac{\partial}{\partial t}\omega$. Whenever the prediction error is larger than its average it leads to a negative $\frac{\partial}{\partial t}\omega$ (negatively valenced affect). Using this in a numerical updating scheme would result in a decrease in estimated precision ω. The opposite would happen when the measured prediction error would be smaller than its average (ω^{-1}) (positively valenced affect). One should be able to see now how Equation 3-1c allows an agent to infer changes in precision and thus infer whether it is deviating or moving towards its homeostatic preferences. That is, when e^2 is larger than the variance it indicates a deviation, when it is smaller than it's average, it indicates moving towards the homeostatic preference. Take note that this can be the result of either prediction errors induced by predictions not matching sensory values or predictions not matching prior expectations (see equation 2-11. It could also be due to sensory or process noise. Take note that the first might be solvable through prediction updates or action, whereas the second is not solvable through this.

One can now see that equation 3-1 determines valence by comparing the measured prediction error to it's average. It is also illustrated in Figure 3-4.

It is essential to note is that every prediction error throughout a hierarchical model is equipped with a precision that can be optimized [5]. The optimization of all these precisions together is what gives rise to affect for the organism.

Take note that everything so far discussed can be assigned "affective consciousness" when it encompasses changes in uncertainty occuring the organisms interoceptive world. The same process in the exteroceptive domain is what Solms would call "perceptual consciousness" or attention.

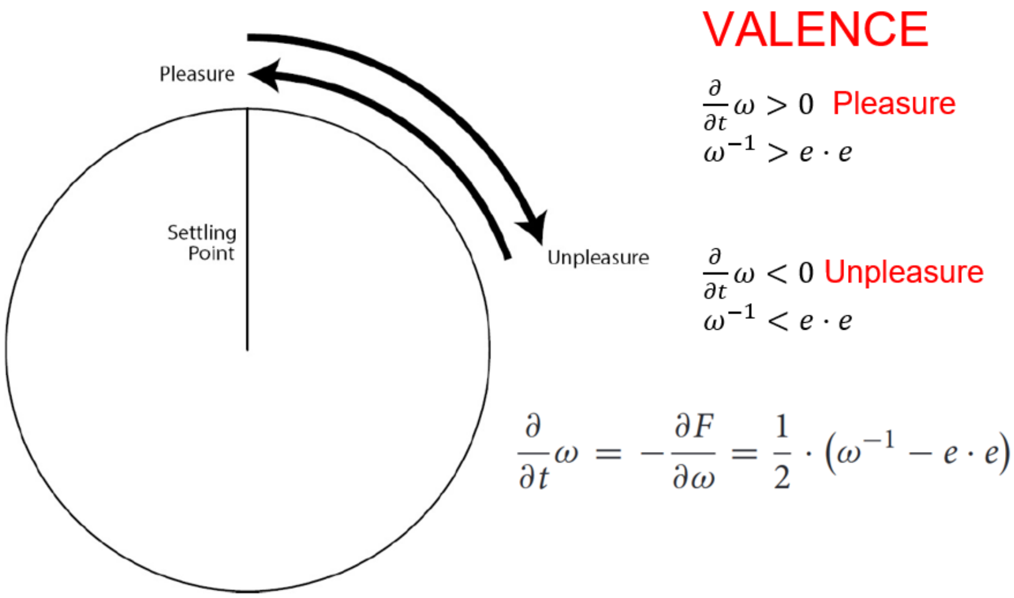


Figure 3-4: Here an illustration on how the precision estimation equation

For obvious biological imperatives, interoceptive "affective consciousness" is prioritised. perceptual consciousness in that sense is subordinate to affect [4]. It can be seen as a contextualisation of affect, where its function is to help maintain the interoceptive homeostatic needs, by appropriately manipulating the external world. It is not just *I feel like this*, but rather *I feel like this about that* [6]. This contextualisation is an essential part of the affective story and will be elaborated on in the next sections as well as in the final model.

3-2-4 Answer to sub-question 1

The process of affect and consciousness revolves around the inference of changes in precision, where increases are registered as positive affect and decreases as negative affect. Here, the crux of Solms' theory on affect as well as of this research lies in Equation 3-1c. It describes the gradient descent with respect to the precision variable ω. This equation optimizes ω, to minimize free energy. This is the part of free energy minimization that Solms associates with affect and consciousness.

The main focus of this research is the computational implementation of this equation in an active inference model, to determine whether it can perform the same functions that affect has in a real organism. This function revolves around the prioritization of different homeostatic needs. This will be further discussed in Section 3-3 and 3-4.

3-3 The function of affect

3-3-1 Introduction and overview

The last section unveiled that affect revolves around the estimation or optimization of precision ω. Here it highlighted that this inference of precision serves as a monitoring mechanism of an organism's ability to maintain its homeostatic preferred equilibria. This determines whether an organsim experiences negatively or positively valenced affect. However, on top of this, precision also functions as a gain modulator determining the influence of prediction errors by regulating the magnitude of the error signal. This underwrites the concept of arousal (magnitude in the vectory analogy). It is an essential property of precision in active inference and plays a key role in the function of affect: the prioritization of needs. This section will highlight this aspect of affect. Doing so, it will provide an answer to the second research sub-question:

- What is the function of affect in both a neurological and computational sense?

It will show that this section will provide a completion upon last section in providing a final definition of affect. This complete definition can then be used to answer the third research sub-question.

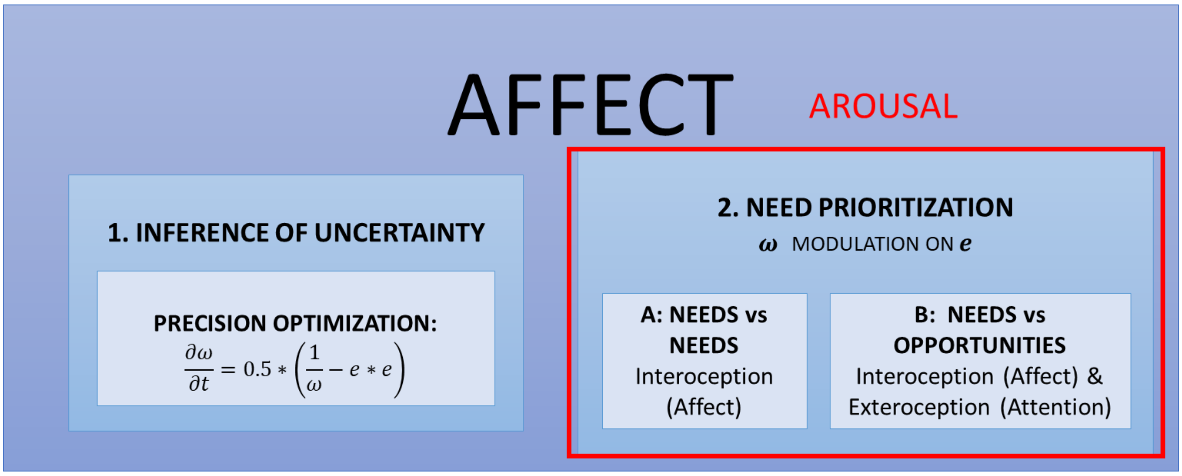


Figure 3-5: The second aspect of affect: Need prioritization. Priortized needs come to affective awareness to the organism as arousal.

The main function of affect as highlighted by Solms in [4, 6, 5] is that it allows the organism to prioritize needs in unpredictable contexts. Which gives the organism an adaptable advantage over non affective types of homeostasis. Computationally, needs are prioritized using precision modulation. This is done in two ways:

(A) The prioritization of needs in relation to other needs.

(B) The prioritization of needs in relation to contextual opportunities and restrictions.

Furthermore, this prioritization using the precisions is defined in [6] using two steps:

1. Needs are prioritized initially by recognizing a specific context and adjusting precisions accordingly.
2. The precisions are then continuously optimised when events unfold.

3-3-2 Why do we feel?

As the generation and experience of affect have been covered in the last section, the question follows as to why this mechanism is beneficial for an organism to have. The answer to this question lies in the improvement of adaptability, especially in unpredictable contexts. In [4] Solms highlights:

> *Feeling enables complex organisms to register—and thereby to regulate and prioritize through thinking and voluntary action—deviations from homeostatic settling points in unpredicted contexts.*

This statement stands central in this research and will be the main focus of this section. As explained in the previous section, affect is not a single variable. Complex biological beings have numerous homeostatic (affective) demands that all need to be satisfied in their own right. This categorization of needs is necessary, as different needs can have different implications for the organism in different contexts. However, this categorization can induce conflict as there are sometimes certain needs that cannot be fulfilled simultaneously. After all, one cannot eat and sleep at the same time. The organism thus needs some way to prioritize its homeostatic needs. Just looking at which prediction error is the largest is not enough, as in some situations a certain need (i.e. prediction error) might be crucial whereas in other situations it is of no importance. [4] And [6] mention that it would in theory be possible for evolution to devise an algorithm that could compute relative survival demands in all predictable situations and thus automatically prioritise actions on this basis. However, this would be a slow model requiring a lot of processing power which is not ideal in life-and-death situations. Even more importantly, such an algorithm would also pose problems when a situation has unpredictable outcomes. As the majority of contexts are however unpredictable to some extent, organisms need a way to conduct themselves appropriately in order to survive.

This is where the role of affect comes in. Affect allows an organism to "feel" its way through the problem. From Section 3-2, it has become clear that deviations from homeostasis are negative and moving towards homeostasis is positive i.e. increasing uncertainty is felt as unpleasure and increasing certainty is felt as pleasure. However, precision does not only determine the valence within a category. It also determines the arousal of a category, which underwrites prioritization of different categories amongst one another. In other words, precision optimization determines which need is felt most saliently at each exact time.

Precision and need prioritization

Taking into account the findings of Section 3-2, the computational implementation of optimizing ω connects well to the need prioritization mechanism as just described. This works as follows: whenever (the square of) the prediction error e of certain modality is larger than expected (ω^{-1}), indicating that the agent is moving into a more uncertain (or less precise) state concerning that modality, it down-regulates the influence of that prediction error as it lowers the precision weighting. At the same time, this would be registered as negatively valenced affect. Whenever the prediction error is smaller than the expected average, the agent is moving into a state of decreasing uncertainty, up-regulating the influence of its prediction errors. This would be registered as positively valenced affect.

As already mentioned there are innumerable homeostatic needs, each with its associated prediction error. It was highlighted that precision acts as a gain modulator on prediction errors meaning precisions determine the weight of the prediction errors in state updates. The prediction error with the highest precision comes forward as arousal. In other words, precision determines which prediction error becomes conscious as felt affect. In short, the changes of precision in the most salient category of need, being the category with the highest expected precision ω in the first place, is what comes on top to the organism as affective awareness (i.e. as arousal) [4] & [5].

Taking in mind the findings of Sections 3-2 and this section. One can thus see how the initial magnitude of ω has a crucial impact in relation to prediction error in two ways: Firstly it determines whether incoming prediction errors induce negatively or positively valenced affect as it sets the threshold for whether precision is increasing or decreasing (see equation 3-1c). Secondly, it also acts as a gain modulator, determining the influence of each prediction error.

Affect is constituted by both these mechanisms. In [4] the following is mentioned:

> *"It is important to note that felt affects typically incorporate both the selected error signal and the ensuing adjustment of cortical (and over longer time frames, subcortical) precisions".*

This summarizes the findings as explained above and reflects figure 3-1.

Two stages of prioritization In [6] there are two stages to the determination of precision and ensuing need prioritzation. First of all, precisions are set by recognizing a specific context and using that to pre-determine which homeostatic affects should be prioritized. The confidence levels and associated baseline precisions here are learnt through experience and set by the brain's memory systems throughout the predictive hierarchy. Neurologically, the prioritization informed by these expected precisions will be performed by a cloud of neuromodulators that are spread throughout the forebrain. This encourages certain channels in the brain whereas others are discouraged. This determines the weight given to current predictions and their errors. In chapters 4 and 5 this process will just be represented by setting the initial precisions of the model.

More interesting in the light of this research, is the second stage of precision optimization. An organism will never be able to perfectly predict precisions, even more so, uncertainties can also change as events unfold. An organism therefore needs to be able to "adjust its precisions on the hoof" [6]. This is where Equation 3-1c comes in. Precisions are recurrently assessed and optimized according to the incoming prediction errors.

From [5, 4] can be quoted : *"changes in subjective quality arise when the amplitude of prediction error changes"*. In other words, the organism "feels" changes in prediction error and thus uncertainty. It adapts the precisions accordingly and thus moves through the environment continuously tailoring the feeling of needs to what is on offer exteroceptively as well as needed interoceptively. This research will mostly be focusing on this continuous adaptation, but it will also slightly touch upon the the effects of setting initial precisions. This will all become more clear in Chapter 5.

3-3-3 Needs vs needs & Needs vs opportunities

Looking further into the prioritization of needs, Solms argues in [4] that an organism determines this prioritization according to a given context:

> *"The prioritization of needs—i.e., the determination as to which need will be felt—must obviously depend crucially upon* ***context (i.e., needs in relation to other needs, and needs in relation to opportunities)****. Feeling is therefore extended onto exteroception (i.e., it is contextualized: "I feel like this about that") and transformed into cognitive consciousness"*
>
> *"This in turn gives rise to voluntary action—and what we loosely call thinking—and, over longer time scales, to learning from experience"*

The first quote is crucial, as it summarizes part of the central requirements of the model that will be explored in the next chapters. It highlights that an organism's prioritization of needs depends on context, where context is a broad term that can manifest in two ways. Firstly context can denote needs in relation to other needs. This basically implies that needs are compared to other needs and thus prioritised in accordance with an organism's interoceptive preferences. That is, which homeostatic need is the most important at a certain time point. This way of going about prioritization of needs is purely focused on interception and thus encompasses "affective consciousness".

Secondly, context can denote needs in relation to opportunities. An organism is very likely to live in a dynamic and unpredictable environment. As a result it might be the case that in some situations the prioritization of a certain need allows for a good opportunity for the organism to reduce it's free energy. However in another situation, prioritization of a different need allows for effective free energy minimization. Solms argues, that because of this, organisms also prioritise needs in accordance with exteroceptive context. As followed from the quote, this way of prioritization requires a connection between intero- and exteroception. This extension of feeling onto exteroception (see quote previous page) thus asks for a combination of interoceptive "affective consciousness" and exteroceptive "perceptual consciousness" or attention. According to [4], in almost every situation exteroception is subordinate to interception, for obvious biological imperatives.

Both of these ways of prioritisation should be included in an affective model.

3-3-4 Adaptive advantage example

Last subsections elaborate on the idea on how "prioritization of needs" offers an adaptative advantage to organisms as opposed to simpler forms of homeostasis. That is, an organism that can adaptively prioritize its needs according to what is available or demanded by the external context.

In the book [6] this is illustrated with a beautiful example:

> *Here is an example I noticed today. When I went for my jog at 7 a.m. it was dark, and when I returned an hour later it was light. (It is winter and I am staying in rural Sussex, writing this book.) Leaving, I passed a field adjacent to the farmhouse where a flock of sheep noticed me and they almost fell over each other to get away. Passing the same field on my return, the same sheep, lying in the same place, barely looked at me. Their startlement in the context of darkness was replaced by boredom in daylight. In short, the context altered the significance of the event 'human running towards me'. At night, this event is prioritised, which snaps the sheep into FEAR mode; by day it is not, and they remain in default-mode SEEKING.*

The crux of this example lies in that the sheep change their affective state and thus behaviour changes as the availability of sensory information provided by the world changes. In other words, the interoceptive needs that are felt by the sheep change upon change of context. This kind of ability can provide an organism with enormous adaptive advantage. By equipping the model that will be presented in the next chapter with the ability to perform the two aforementioned tasks, the model that will be described will demonstrate this improved adaptability.

3-3-5 Answer to sub-question 2

To summarize, by optimizing precisions, the organism can determine which prediction error i.e. which homeostatic need, is prioritized. In more psychological terms, this can be described as an organism "feeling" its deviations from homeostasis by inferring the associated precisions and using this precision to make the prediction error come forward as arousal.

Here, increases in precision will be registered as positive and decreases will be registered as negative. This in turn leads to a series of unfolding choices in an expected context, guided by expected precisions [6]. Crucial here is the distinction between prioritization of needs amongst one another versus the prioritization of needs in relation to exteroceptive context This answers the second sub-question.

3-4 Conclusion and final model requirements

It has been explained in Section 3-2 that affects arise through deviations from homeostatic equilibria. The feeling of these affects arises from the inference of the uncertainty about the deviations. Computationally this is represented by precision optimization. Section 3-3 then elaborated on this by explaining the function of affect in prioritising needs. This is formally represented by modulation through precision ω. Section 3-3 elaborated on this need prioritization, by highlighting the two ways an organism can go about this prioritization according to the affective story: 1. needs in relation to other needs. 2. The relation of needs with respect to opportunities as represented by the exterior.

- What simulation results are needed to prove the model works according to Solms definition?

To accommodate for a model and simulation that applies to Solms' description of affect, as presented in sections 3-2 and 3-3, there are two major points and concepts that need to be present in the model. First and foremost, as affect is computationally defined by Solms as the precision updating scheme as presented in Equation 3-1c, this needs to be included in the model. This is one of the two crucial aspects of affect. However, to assess whether this precision updating scheme is working in a manner that complies with Solms definition not only of affect itself but also with Solms definition of the function of affect, the prioritization of needs, the model needs to be constructed in such a way that it can experimentally demonstrate this. As concluded from the previous chapter, this prioritization is done by using the precision ω as a gain on prediction errors. Furthermore, it has been clarified that there are two ways in which an organism can go about this. This is all again illustrated in Figure 3-6.

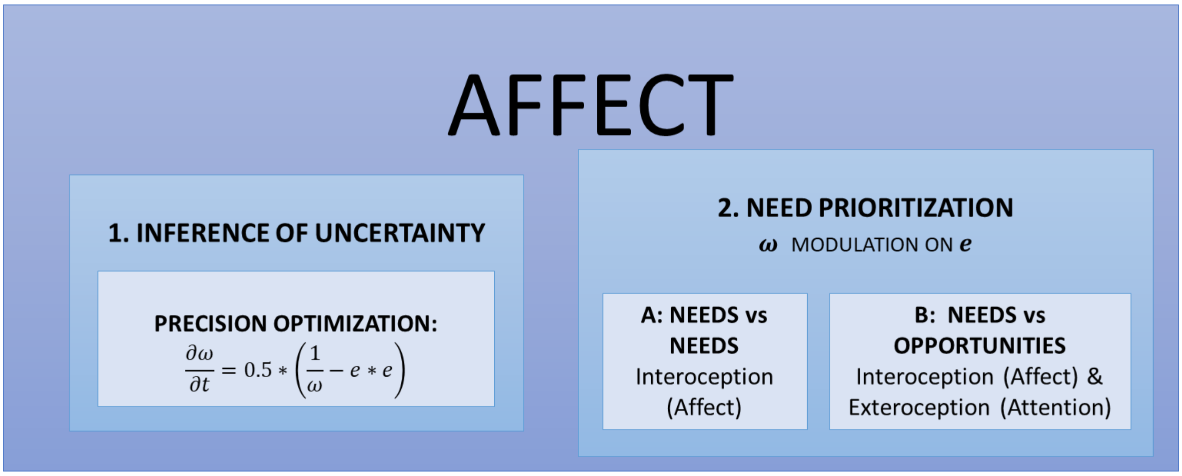


Figure 3-6: Definition of affect

Thus the final requirements of simulation results of an affective model need to include the following points:

1. Inference of uncertainty through precision optimization.

2. Need prioritization, as follows:
   (A) The prioritization of needs in relation to other needs.
   (B) The prioritization of needs in relation to contextual opportunities and restrictions.

These points set up the requirements for the implementation of affect in active inference that is used in this research. Chapter 4 will elaborate on the specific model design that is going to be used to achieve this. Chapter 5 will provide two simulational experiments that successfully implement this model and illustrate that these two prioritisations can be performed. This list and Figure 3-6 will be used as a reference in the next chapters to assist in the explanation of the affective parts of the model and simulation results that will follow from the next chapters.

Chapter 4

Affective active inference model

4-1 Introduction

This chapter is aimed at providing an affective active inference model, accompanied by a thorough mathematical explanation, but more importantly an explanation of model choices and how they relate to Solms' insights that were discussed in Chapter 3. The model provides a possible implementation of Mark Solms' notion of affect in active inference.

4-1-1 Introduction to Hydar

The model that is going to be illustrated in this chapter is a simplification of Hydra vulgaris [15]. Hydra is a small freshwater animal that belongs to the group of Cnidarians, which are contemporary representatives of some of the earliest animals in evolution to have a nervous system. In [16], it is highlighted that this nervous system is a precursor of what today would be the brainstem. Taking this into account, a computational simulation representing Hydra would be suitable for performing the experiments on affect in active inference. This implementation is going to be called Hydar.

The simplicity of Hydra's neural system allowed the exploration of structural and functional design principles of neural circuits. In [15], it was found that Hydra contained multiple, anatomically non-overlapping functional neural networks. Each of those is associated with specific behaviours (see figures 4-1 and 4-2). This compartmentalization of networks provides an inspiration that is very suitable for this experiment. As has already been touched upon, the simulations provided in this research are to prove that by using affect and precision optimization it is possible to prioritize homeostatic needs.

As will be shown in this and the following chapter, this prioritization mechanism consists of two separate modules that dictate two different behaviours. These two modules, combined with an action module could represent a structure similar to the separation of neural networks and associated behaviours as depicted in [15] and Figure 4-1 and 4-2. Note however that the

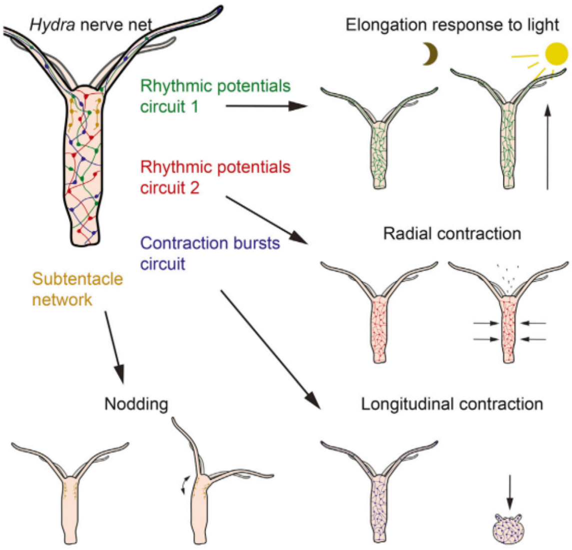


Figure 4-1: The nervous system of Hydra is traditionally described as made of two nerve nets. By using calcium imaging, Dupre and Yuste demonstrate the existence of multiple circuits within these nerve nets and show with which behaviour they are associated

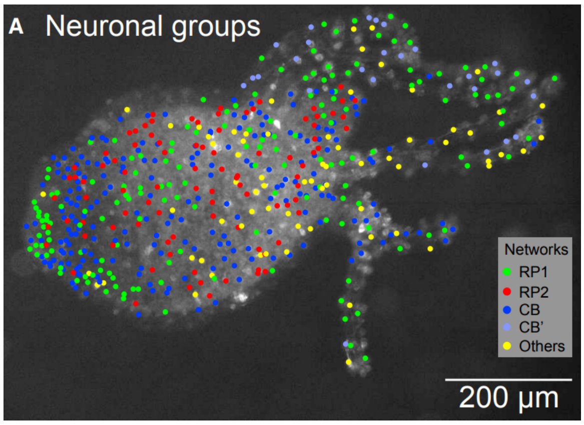


Figure 4-2: Topographical distribution of neurons in Hydra (same dataset as Figure 1D), grouped into five categories: rhythmic potential 1 (RP1; green), rhythmic potential 2 (RP2; red), longitudinal CBs (dark and light blue), and other neurons (others; yellow). CB0 indicates neurons of the tentacles that did not fire during the two CB events of this time window but fired during another CB event

simulation will be merely inspired by Hydra and its separation of networks and not contain the different networks as depicted in figures 4-1 and 4-2 (these images and captions are directly taken from [15]). The two types of behaviour that are going to be performed by the model moving towards the surface (0m) or moving towards a depth of (2m). These two behaviours are dictated by Hydar's temperature and food preferences respectively. The computational representation of Hydra vulgaris will be done in Python. Before diving more deeply into the model design, it should be acknowledged that the inspiration for using the Hydar simulation is taken from [17].

4-1-2 Model requirements

Having introduced Hydar, this section will look more closely into its overall structure. The goal of the Hydar simulation is to show that through the optimization of precision, need prioritization within an agent can take place. As concluded from Section 3-4 the requirements for what the model needs to be able to perform are as follows:

1. Inference of uncertainty through precision optimization.
2. Need prioritization, as follows:
   (A) The prioritization of needs in relation to other needs.
   (B) The prioritization of needs in relation to contextual opportunities and restrictions.

Whereas Chapter 5 will illustrate how the devised model successfully performs these two tasks, the focus of this chapter lies more in explaining the model choices in relation to these two ways of prioritisation. To properly account for and thus illustrate both of these mechanisms, four major design requirements need to be included in the model architecture.

First of all, to be able to prioritize needs (whether it is in relation to other needs or in relation to context), the model has to include **multiple needs** in the first place. Secondly, as homeostatic needs are inherently about selfhood, Hydar needs a way to monitor its own internal states using an **interoceptive system**. Thirdly, as the model needs to prioritize needs through contextualization (more specifically contextual opportunities), the system needs a way to infer its external world using an **exteroceptive system**. Lastly, Hydar needs **action** in order to fulfil its prioritised needs.

For clarity, these four requirements are again listed below:

1. Hydar needs **multiple** interoceptive needs that can compete.
2. Hydar needs an **Interoceptive** system that can measure and prioritize interoceptive prediction errors.
3. Hydar needs **Exteroception** to prioritize these needs according to context.
4. Hydar needs **action** to perform according to this prioritization.

To address these requisites, Hydar is equipped with two interoceptive sensors. One of these sensors gauges internal temperature, while the other assesses hunger levels. To satisfy its homeostatic demands, Hydar can manoeuvre both upward and downward in varying depths. Additionally, Hydar features two exteroceptive sensors: one for temperature and the other for detecting food. Utilizing these sensors, Hydar can deduce its positioning in relation to these external variables. Given that one of the specifications for Hydar's model involves contextualizing interoception with exteroception, it necessitates a mechanism to connect the two. This integration is achieved through a hierarchical layer that communicates with and receives input from both its interoceptive and exteroceptive modules. The hierarchical layer encompasses priors for both temperature and food modalities. These priors establish Hydar's desired values for each modality. Hydar can be divided into two modules that try to regulate their specific modality: food and temperature. Each of the two modules has an interoceptive sub-module, an exteroceptive part and a hierarchical part, where each of those sub-modules has its own prediction error and each of those sub-modules has its own precision. The exteroceptive parts of each modality have a shared influence on action. The scheme just described is illustrated in a simple manner in Figure 4-3

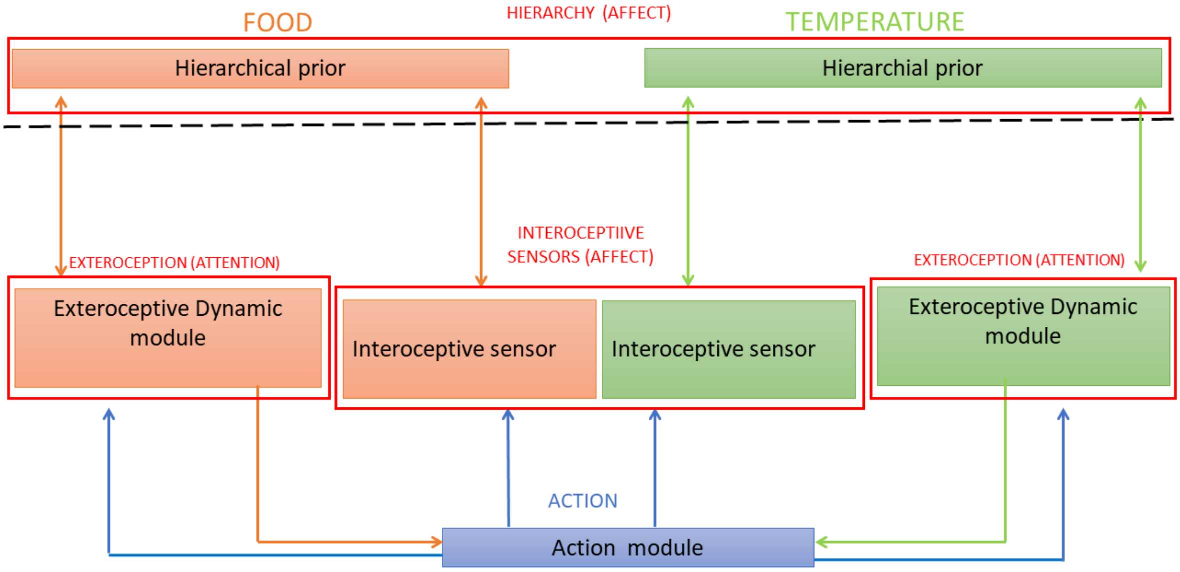


Figure 4-3: Overview of Hydar's structure. In green the temperature module and in orange the food module. Each module contains 3 sub-modules. The interoceptive sensor and hierarchical prior are both associated with the estimation of the interoceptie state and therefore associated with "affective consciousness". The exteroceptive modules are associated with "perceptual consciousness".

4-1-3 What is affect in this model?

As has been thoroughly explained in Chapter 3, affect is associated with two concepts: precision optimization and need prioritization, which is, as in Chapter 3, illustrated in the figure below:

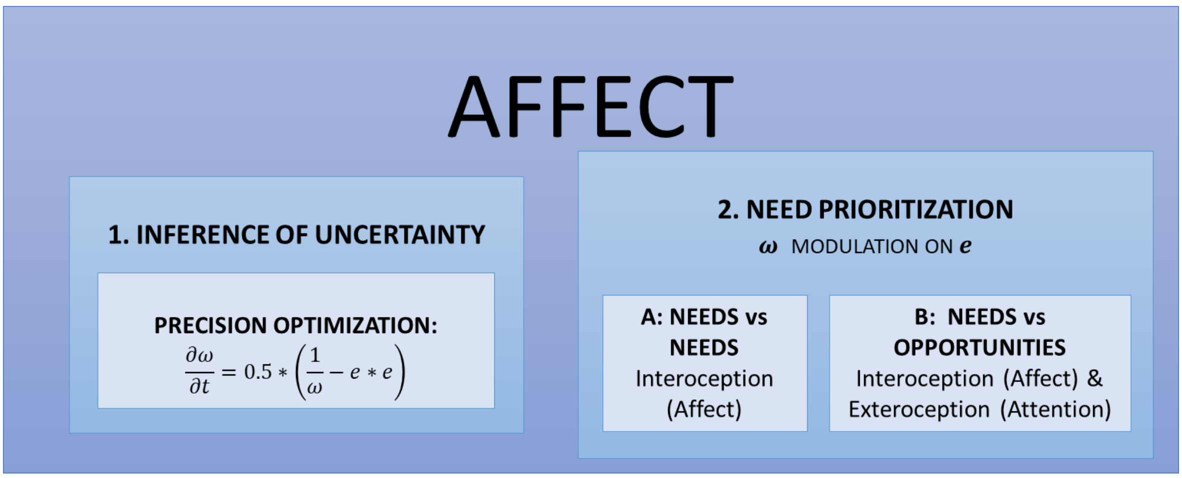


Figure 4-4: First part of affect: The inference of precision

1. Inference of precision For hydar to survive, it needs to ensure the homeostatic needs of its internal temperature and food. Both are measured by its interoceptive system. Deviations from the homeostatic need are valanced positive if the precision increases (less variance, less uncertainty) and negative if the precision decreases (more variance, more uncertainty).

Note that, as defined by Solms, 'affective consciousness' is associated with the inference of interoceptive precision and "perceptual consciousness" or attention is associated with exteroceptive precision.

In the case of the model described in Figure 4-3, the precisions that are associated with interoception are the precision of the interoceptive sensors, as well as the precision of the hierarchical priors. As will become more clear in the coming sections, these precisions all influence the internal state estimations in both the food and the temperature modalities. Affect as a whole is thus represented by the inference of precision in the internal sensory sub-modules as well as the hierarchical sub-modules. Precision estimation of these modules determine whether the prediction errors are registered as positively or negatively valenced affect.

The same mechanism is included in the exteroceptive sub-modules. These are however meant to infer exteroceptive precision (sensory noise). They have the function to modulate which sensors should be used and which ones shouldn't. This is called "perceptual consciousness" or attention. The model can use this to asses and thus act upon external opportunities and restrictions. See Figure 4-5 for illustration.

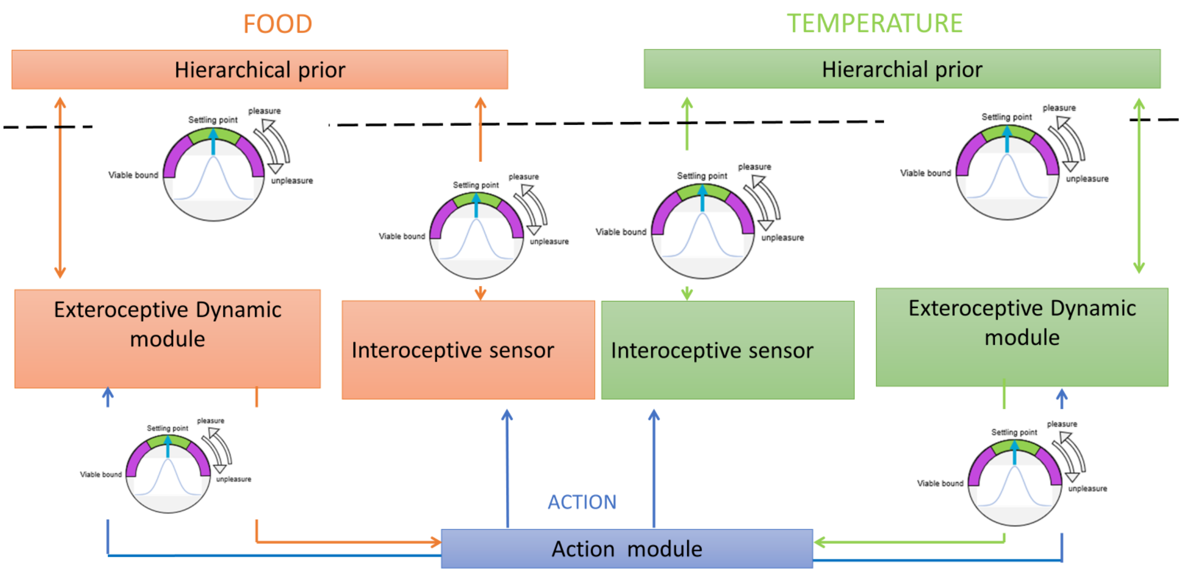


Figure 4-5: This figure illustrates that every sub-module with a precision that can be estimated using a prediction error. Thus every sub-module has the ability to infer changes in precision and thus reflect valence in its own specific category. Here the interoceptive sensors and hierarchical priors are associated with interoceptive state estimation and thus affect. The exteroceptive modules are associated with attention

2. Prediction error prioritization Secondly affect includes the feeling of prioritized prediction errors (which are an indication of uncertainty). This part of affect is depicted in the second part of Figure 4-4. Need prioritization determines which need (or prediction error) is felt. It is determined by either comparing needs to other needs or comparing needs to exteroceptive context, as shown in Figure 4-4. The prioritized need drives the behaviour of the active inference system. E.g. when hunger is the prioritized need, Hydar will want to go to the ideal food depth. The prioritized need is the result of the dynamical active inference calculations and not a single parameter that can be pointed out in the system.

In more technical terms, the prioritised prediction errors, i.e. prediction errors with a higher precision are going to come forward to Hydars affective awareness. Note that this holds true for both affect in the interoceptive and hierarchical sub-modules and attention in the exteroceptive sub-modules. The crux of this idea is that Hydar will act upon the prediction errors with the highest precisions as if it is "consciously aware" of or "aroused" by these errors.

Crucial to take into account is that in a more sophisticated model, that could more realistically simulate a real organism there would be a much deeper hierarchy with many more layers, each with their own fluctuating, optimizable precisions and thus affective values. The totality of these fluctuating precisions would represent the activity of affective consciousness. Here prediction errors in this hierarchy with the highest afforded precision would come to "affective awareness" to the agent and thus become present as "conscious thought".

The next sections will be devoted to highlighting the model in more detail. The end of Section 4-4, will conclude by getting back to the ideas presented above, using the detailed description of the model. See Figure 4-6 for illustration.

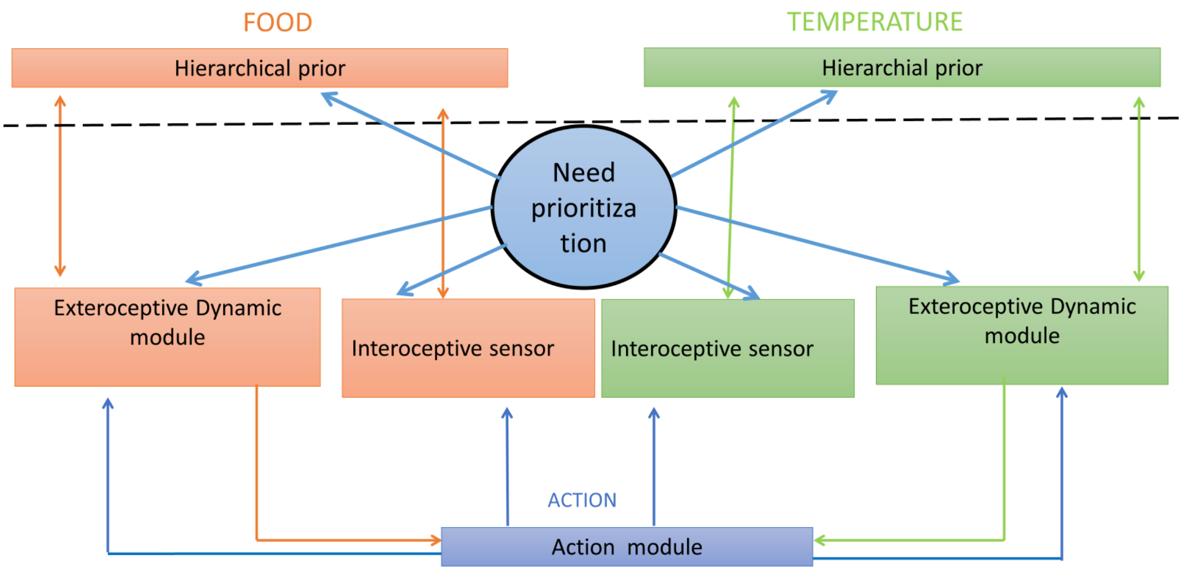


Figure 4-6: Need prioritization illustrated in the Hydar model. As every sub-module is equipped with a precision it can be subject to need prioritization.

4-2 Generative process and Generative model

Firstly the Generative process and the Generative Model of the Hydar simulation will be described. This will then be followed by a detailed summary of the hidden state estimations through gradient descent. Lastly, the precision updating schemes will be discussed.

4-2-1 Model equations

Hydar's external and internal world simplifications

Hydar lives in a world where a number of characteristics are simplified. This has been done to make sure that the results can portray Hydar's need prioritization as clearly as possible.

To start with, Hydar lives in a world where temperature and foodvalues are identical to the value of the depth x. In other words, there will be a 1 to 1 mapping between the depth, temperature and Food. This means that at a depth of 1, Hydar's exteroceptive temperature sensor should be giving a temperature reading of 1 temperature unit and Hydar's interoceptive temperature sensor should be giving a reading of 1 foodunit. Take note that these readings can potentially vary due to added noise. Also be aware that even though they have the same mapping from x, the foodand temperature readings are two separate signals, originating from their own sensors.

This 1 to 1 mapping is also applied to the interoceptive sensors. These do not receive any noise. In other words, Hydar's internal temperature reading and his internal foodreading are equal to its depth.

As will be shown shortly, the preferred values as specified by the priors, also conform to this 1 to 1 mapping. That is, Hydar's prior preferences represent and specify its preferred foodand temperature reading, however, due to the mapping this will also be translated to a preferred depth that is equal to those values.

All of this should become more clear in the following paragraphs.

Generative Process

The generative process is shown in the equations below:

$$\begin{aligned} x &= depth \\ \dot{x} &= f_{gp}(x) = u \\ \tilde{g}_{gp}(\tilde{x})_{food} &= \tilde{x} = \begin{bmatrix} x \\ \dot{x} \end{bmatrix} \\ \tilde{g}_{gp}(\tilde{x})_{temp} &= \tilde{x} = \begin{bmatrix} x \\ \dot{x} \end{bmatrix} \end{aligned} \tag{4-1}$$

Here $\dot{x}$ is the time derivative of the depth and is equated to the action variable u. This simply means that Hydar can set its own speed. Furthermore, $g_{gp}(x)_{temp} = x$ and $g_{gp}(x)_{temp} = x$

denote that in this world, the depth value directly determines the value for temperature and foodand thus will always be equal. This is due to the 1 to 1 mapping as highlighted above. Take note that when sampling these variables using its sensors, there might still be a discrepancy between the sensory value and the real depth due to additional noise:

$$\begin{aligned} \tilde{y}_{temp} &= \tilde{g}_{gp}(\tilde{x})_{temp} + \tilde{z}_{temp} \\ \tilde{y}_{food} &= \tilde{g}_{gp}(\tilde{x})_{food} + \tilde{z}_{food} \end{aligned} \tag{4-2}$$

In other words, when noise is added to the process, Hydar's sensory values will fluctuate around the real depth x (and thus the real foodand real temperature values).

For the purpose of this model, the internal sensory states φ do not contain any noise and are, due to the 1 to 1 mapping, also identical to the actual depth x. What this means is that Hydar's internal temperature will change immediately with the change of position. The same goes for its interoceptive food(or hunger):

$$\begin{aligned} \varphi_{temp} &= x \\ \varphi_{food} &= x \end{aligned} \tag{4-3}$$

Hydar's priors are defined as follows:

$$\begin{aligned} \mu_{P_{temp}} &= 0 \\ \mu_{P_{food}} &= 2 \end{aligned} \tag{4-4}$$

Remember that these values denote preferred interoceptive states and not depth directly. The associated preferred depths will however be valuated the same due to the 1 to 1 mapping.

Generative Model

As mentioned in Section 4-1 the model needs an interoceptive as well as an exteroceptive part.

Exteroceptive For the exteroceptive part a dynamic model is used.

$$\begin{aligned} D\tilde{\mu}_{x_{temp}} &= \tilde{f}_{gm}(\tilde{\mu}_{x_{temp}}, \mu_{temp}) + \tilde{w}_{temp} \\ \tilde{y}_{temp} &= \tilde{g}_{gm}(\tilde{\mu}_{x_{temp}}) + \tilde{z}_{temp} \\ D\tilde{\mu}_{x_{food}} &= \tilde{f}_{gm}(\tilde{\mu}_{x_{food}}, \tilde{\mu}_{food}) + \tilde{w}_{food} \\ \tilde{y}_{food} &= \tilde{g}_{gm}(\tilde{\mu}_{x_{food}}) + \tilde{z}_{food} \end{aligned} \tag{4-5}$$

Here the "g" functions (see Chapter 2) are defined according to the 1 to 1 mapping and the "f" functions are defined as a simple attractor state dynamic:

$$\begin{aligned}
\tilde{f}(\tilde{\mu}_{x_{temp}}, \mu_{temp}) = -\tilde{\mu}_{x_{temp}} + \tilde{\mu}_{temp} &= \begin{bmatrix} -\mu_{x_{temp}} + \mu_{temp} \\ -\dot{\mu}_{x_{temp}} \end{bmatrix} \\
\tilde{g}_{gm}(\tilde{\mu}_{x_{temp}}) = \tilde{\mu}_{x_{temp}} &= \begin{bmatrix} \mu_{x_{temp}} \\ \dot{\mu}_{x_{temp}} \end{bmatrix} \\
\tilde{f}(\tilde{\mu}_{x_{food}}, \tilde{\mu}_{food}) = -\tilde{\mu}_{x_{food}} + \tilde{\mu}_{food} &= \begin{bmatrix} -\mu_{x_{food}} + \mu_{food} \\ -\dot{\mu}_{x_{food}} \end{bmatrix} \\
\tilde{g}_{gm}(\tilde{\mu}_{x_{food}}) = \tilde{\mu}_{x_{food}} &= \begin{bmatrix} \mu_{x_{food}} \\ \dot{\mu}_{x_{food}} \end{bmatrix}
\end{aligned} \tag{4-6}$$

Here $\tilde{\mu}_{x_{temp}}$ describes Hydar's estimate of the external temperature and $\tilde{\mu}_{x_{food}}$ describes Hydar's estimate of the external Food.

μ_{food} And μ_{temp} are the estimates of the internal temperature and foodstates. These are special as they have two roles. They first of all act as an estimate of the internal foodand temperature values, this will become clear shortly. Secondly, as shown above, they act as an attractor state (or causal state in [8])in the "f" functions. In other words, μ_{food} and μ_{temp} determine where Hydar wants to go using its exteroceptive modules. Hydar only takes into account the 0th order internal estimates. In other words, it does not use generalised coördiates. Thus they are defined as follows:

$$\begin{aligned}
\tilde{\mu}_{temp} &= \begin{bmatrix} \mu_{temp} \\ 0 \end{bmatrix} \\
\tilde{\mu}_{food} &= \begin{bmatrix} \mu_{food} \\ 0 \end{bmatrix}
\end{aligned} \tag{4-7}$$

Note that this vector notation is used such that Equation 4-7 can be fitted in Equation 4-6

Estimating the dynamics of this variable is done using the following prediction errors:

$$\begin{aligned}
\tilde{e}_{x_{temp}} &= D\tilde{\mu}_{x_{temp}} - \tilde{f}\left(\tilde{\mu}_{x_{temp}}, \tilde{\mu}_{temp}\right) \\
\tilde{e}_{x_{food}} &= D\tilde{\mu}_{x_{food}} - \tilde{f}\left(\tilde{\mu}_{x_{food}}, \tilde{\mu}_{food}\right)
\end{aligned} \tag{4-8}$$

Furthermore, the estimates of the external $\tilde{\mu}_x$ states are compared to the sensory states $\tilde{y}$ leading to the exteroceptive prediction errors. The noise terms $\tilde{z}_{temp}$ and $\tilde{z}_{food}$ in Equation 4-5 make up for these prediction errors, evaluated as discrepancies between the sensory signals $\tilde{y}$ and the prediction made by the model with the g function. These discrepancies can consist of a mismatch between a predicted value and a measured value due to a wrong prediction, as well as noise that had been added to the sensor.

$$\begin{aligned}
\tilde{e}_{y_{temp}} &= \tilde{y}_{temp} - \tilde{\mu}_{x_{temp}} \\
\tilde{e}_{y_{food}} &= \tilde{y}_{food} - \tilde{\mu}_{x_{food}}
\end{aligned} \tag{4-9}$$

For the sake of readability, the collection of both external states estimates $\mu_{x_{temp}}$ and $\mu_{x_{food}}$ state estimates can be written as μ_x (or $\tilde{\mu}_x$ in the generalized notation).

Interoceptive The interoceptive part only consists of the sensory readings φ_{food} and φ_{temp} as described in Equation 4-3. These readings are compared to the internal state estimates μ_{temp} and μ_{food} resulting in interoceptive prediction errors.

$$\begin{aligned} \varepsilon_1 &= \varphi_{temp} - \mu_{temp} \\ \varepsilon_2 &= \varphi_{food} - \mu_{food} \end{aligned} \tag{4-10}$$

Note that for the sake of readability, in some cases the collection of both internal state estimates can be written as μ. In this case, μ refers to the internal state estimates in their own modality. Similarly, the external state estimates from the previous paragraph will occasionally be compactly denoted as μ_x

Hierarchical layer As just mentioned, the interoceptive and exteroceptive parts are connected through a hierarchical layer. This hierarchical layer compares internal states $\tilde{\mu}_{temp}$ and $\tilde{\mu}_{food}$ to a prior value.

$$\begin{aligned} \varepsilon_3 &= \mu_{temp} - \mu_{P_{temp}} \\ \varepsilon_4 &= \mu_{food} - \mu_{P_{food}} \end{aligned} \tag{4-11}$$

The value of this prior determines the preferred internal state of the Hydar. This hierarchical layer is connected to the interoceptive sensory modules similarly to Equation 2-12:

$$\begin{aligned} \mu_{food} &= \mu_{P_{food}} + z_{\varepsilon_4} \quad & \mu_{temp} &= \mu_{P_{temp}} + z_{\varepsilon_3} \\ \varphi_{food} &= \mu_{food} + z_{\varepsilon_2} \quad & \varphi_{temp} &= \mu_{temp} + z_{\varepsilon_1} \end{aligned} \tag{4-12}$$

Note that, even though there is no noise added to the internal sensors as mentioned above (and illustrated in Equation 4-3), there is still a z term in every equation. These terms, account for the prediction errors ε from Equations 4-10 and 4-11 which are a result of the discrepancies between the internal state μ and the prior μ_P and between internal state μ and the internal sensory value φ.

It can be seen how the internal state μ connects the interoceptive layer with the layer above. On top of this, as μ is used in the exteroceptive state dynamics as an attractor state, Hydar's interoception is connected to its exteroception. We can use equation a combination of Equations 2-12 and 2-14 to form a scheme like below:

$$\begin{aligned} \mu_{food} &= \mu_{P_{food}} + z_{\varepsilon_4} \quad & \mu_{temp} &= \mu_{P_{temp}} + z_{\varepsilon_3} \\ D\tilde{\mu}_{x_{food}} &= \tilde{f}(\tilde{\mu}_{x_{food}}, \tilde{\mu}_{food}) + \tilde{w}_{food} \quad & D\tilde{\mu}_{x_{temp}} &= \tilde{f}(\tilde{\mu}_{x_{temp}}, \tilde{\mu}_{temp}) + \tilde{w}_{temp} \\ \tilde{y}_{food} &= \tilde{g}_{gm}(\tilde{\mu}_{x_{food}}) + \tilde{z}_{food} \quad & \tilde{y}_{temp} &= \tilde{g}_{gm}(\tilde{\mu}_{x_{temp}}) + z_{\tilde{temp}} \end{aligned} \tag{4-13}$$

4-2-2 Back to Hydar's model architecture

Figure 4-7 illustrates how Equation 4-13 fits into the overall architecture of the model. Here each block contains its respective part of the Generative Model as well as the prediction error

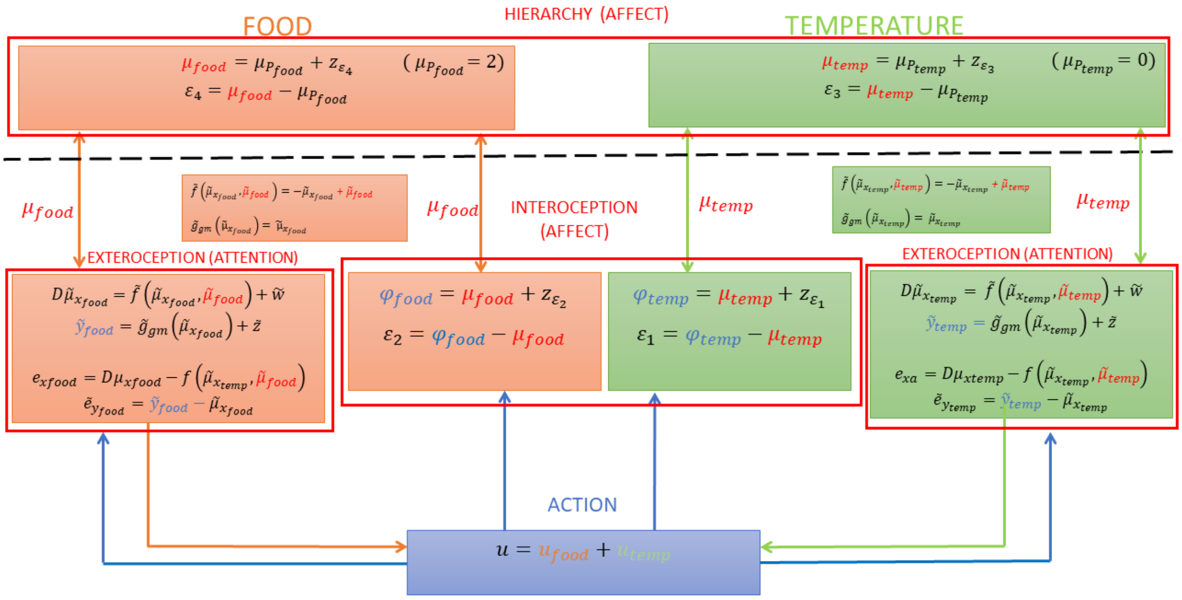


Figure 4-7: This figure shows Hydar's generative model slotted into the model's architecture. Take note of μ_{food} and μ_{temp} being highlighted in red to illustrate connections between the two hierarchical layers. The variables highlighted in blue are the sensory states and are all directly influenced by action.

that belongs to that part of the model, as described in Equations 4-5- 4-12. There are a few things that should be taken note of.

First of all, take into account that the temperature and foodmodules themselves are completely separate and independently influence action.

Secondly, note that the internal state estimates μ_{food} and μ_{temp} (together μ) are highlighted in red all across the figure. This is to emphasise their role in connecting the three different blocks in each module. Or more precisely, the interoceptive submodules and hierarchical submodules together make an evaluation of μ. This μ is then used as an attractor state in the exteroceptive sub-modules.

Comprehending this structure is essential to understanding Hydar's behaviour and later experiments. Central in this argument is that when estimating this internal state μ (through a gradient descent, as will be clarified in the next subsection), the information, or more precisely the prediction errors, provided by all three blocks is taken into account. As Hydar will be looking to minimize free energy, the estimation of μ will be some sort of precision-weighted average of the information processed in the three blocks, with ideally the smallest prediction error possible. Here μ is compared to the interoceptive sensory states φ (ε_1 and ε_2) as well as to the prior preferences μ_P (ε_3 and ε_4). Furthermore, the μ's are also used in the external state dynamics errors e_x. The gradient descent is to update μ such that these prediction errors are minimized.

Regarding the exteroceptive state dynamics, μ can be seen as an attractor state (or differently called causal state) in the f functions of the exteroceptive blocks. Hydar performing a gradient descent on $\tilde{\mu}_x$ defines its movement towards this attractor state μ which in turn, determines action, which is Hydar's second tool to minimize prediction error. Action in turn closes the loop and influences the sensory states which are then used again to estimate the state estimations. This will all be further elaborated on in the next sections after state estimation and precision estimation with gradient descent are covered. As for now, it can be of value to keep this in mind.

Lastly, all sensory variables are highlighted in blue to indicate that they are directly influenced by the action Input.

4-3 Gradient descents

4-3-1 Introduction

The last section describes Hydar's world or generative process, its internal generative model that he uses to make predictions about this world and the last section illustrates how different parts of his model are organised, which parts are connected and what parts are separate. This section will cover the gradient descents used to determine the values of the variables $\mu_{x_{food}}$ and $\mu_{x_{temp}}$, μ_{food} and μ_{temp} and the action input u. As explained in Chapter 2, the variables can be determined by taking a gradient descent of free energy with respect to that variable. This section will define the free energy equation followed by the derivations that are needed to determine the required variables.

4-3-2 Free Energy equation and gradient descent derivations

Free Energy equation Having defined the generative model in the last section, it is possible to set up the free energy Equation as follows:

$$
\begin{aligned}
F = \frac{1}{2}(\tilde{e}^T_{x_{food}}\tilde{\prod}_w \tilde{e}_{x_{food}} + \tilde{e}^T_{x_{temp}}\tilde{\prod}_w \tilde{e}_{x_{temp}} + \tilde{e}^T_{y_{food}}\tilde{\prod}_{z_{food}} \tilde{e}_{y_{food}} + \tilde{e}^T_{y_{temp}}\tilde{\prod}_{z_{temp}} \tilde{e}_{y_{temp}} \\
+\varepsilon_1\omega_1\varepsilon_1 + \varepsilon_2\omega_2\varepsilon_2 + \varepsilon_3\omega_3\varepsilon_3 + \varepsilon_4\omega_4\varepsilon_4 \\
-2ln\tilde{\prod}_w - ln\tilde{\prod}_{z_{food}} - ln\tilde{\prod}_{z_{temp}} - ln(\omega_1) - ln(\omega_2) - ln(\omega_3) - ln(\omega_4))
\end{aligned}
\tag{4-14}
$$

To reduce the free energy, it is necessary to compute the derivatives with respect to the mentioned variables. The following paragraphs will outline these derivative calculations. The prediction errors, as defined by Equations 4-8, 4-9, 4-10, and 4-11, play a pivotal role in this process. Because of the quadratic terms present in the free energy equation, the variables included in each prediction error determine which of those terms are considered in each derivative. This can be seen in the following derivations.

Gradient Descent on μ

The estimates of the internal states μ_{food} and μ_{temp} play a crucial role in Hydar's functioning. Figure4-7 shows that they are used in all three parts of both modules. As explained the μ's is a weighted average of their respective blocks. To determine the optimal values, Hydar will be looking to minimize free energy by minimizing the prediction errors associated with μ_{food} and μ_{temp}. This will be clarified by inspecting the gradient descent equations below:

$$
\begin{aligned}
\frac{\partial F}{\partial \mu_{food}} &= \frac{\partial \tilde{e}_{x_{food}}}{\partial \mu_{food}}^T \tilde{\prod}_w \tilde{e}_{x_{food}} + \frac{\partial \varepsilon_2}{\partial \mu_{food}}\omega_2\varepsilon_2 + \frac{\partial \varepsilon_4}{\partial \mu_{food}}\omega_4\varepsilon_4 \\
\frac{\partial F}{\partial \mu_{\ temp}} &= \frac{\partial \tilde{e}_{x_{temp}}}{\partial \mu_{temp}}^T \tilde{\prod}_w \tilde{e}_{x_{temp}} + \frac{\partial \varepsilon_1}{\partial \tilde{\mu}_{temp}}\omega_1\varepsilon_1 + \frac{\partial \varepsilon_3}{\partial \tilde{\mu}_{temp}}\omega_3\varepsilon_3
\end{aligned}
\tag{4-15}
$$

Taking into account and filling in Equations 4-6, 4-8, 4-9, 4-10, and 4-11 into Equation 4-15, where Equation 4-8, comprises of vectors of p = 2 such that:

$$\begin{aligned}\tilde{e}_{xfood} &= \begin{bmatrix}\dot{\mu}_{x_{food}} \\ 0\end{bmatrix} - \begin{bmatrix}-\mu_{xfood} + \mu_{food} \\ -\dot{\mu}_{x_{food}}\end{bmatrix} \\ \tilde{e}_{xtemp} &= \begin{bmatrix}\dot{\mu}_{x_{temp}} \\ 0\end{bmatrix} - \begin{bmatrix}-\mu_{xtemp} + \mu_{temp} \\ -\dot{\mu}_{x_{temp}}\end{bmatrix}\end{aligned} \tag{4-16}$$

Equation 4-15 leads to Equation 4-17.

$$\begin{aligned}\frac{\partial F}{\partial \mu_{food}} &= \begin{bmatrix}-1 & 0\end{bmatrix} \tilde{\prod}_w \tilde{e}_{x_{food}} - \omega_2\varepsilon_2 + \omega_4\varepsilon_4 \\ \frac{\partial F}{\partial \mu_{temp}} &= \begin{bmatrix}-1 & 0\end{bmatrix} \tilde{\prod}_w \tilde{e}_{x_{temp}} - \omega_1\varepsilon_1 + \omega_3\varepsilon_3\end{aligned} \tag{4-17}$$

These derivatives can finally be formulated like below such that they can be used in a Gradiënt descent:

$$\begin{aligned}\dot{\mu}_{food} &= -\frac{\partial F}{\partial \mu_{food}} \\ \dot{\mu}_{temp} &= -\frac{\partial F}{\partial \mu_{temp}}\end{aligned} \tag{4-18}$$

To sum up, Equations 4-15- 4-18 are used by Hydar to determine the estimates of the internal states μ_{food} and μ_{temp}.

Gradient descent on $\tilde{\mu}_x$

The second set of variables that Hydar will perform a Gradiënt Descent on is the estimates of the external states. As explained in the section above, $\tilde{\mu}_x$ is pulled towards the internal state μ.

$$\begin{aligned}\frac{\partial F}{\partial \tilde{\mu}_{x_{food}}} &= \frac{\partial \tilde{e}_{x_{food}}}{\partial \tilde{\mu}_{x_{food}}}^T \tilde{\prod}_w \tilde{e}_{x_{food}} + \frac{\partial \tilde{e}_{y_{food}}}{\partial \tilde{\mu}_{x_{food}}} \tilde{\prod}_{z_{food}} \tilde{e}_{y_{food}} \\ \frac{\partial F}{\partial \tilde{\mu}_{x_{temp}}} &= \frac{\partial \tilde{e}_{x_{temp}}}{\partial \tilde{\mu}_{x_{temp}}}^T \tilde{\prod}_w \tilde{e}_{x_{temp}} + \frac{\partial \tilde{e}_{y_{temp}}}{\partial \tilde{\mu}_{x_{temp}}} \tilde{\prod}_{z_{temp}} \tilde{e}_{y_{temp}}\end{aligned} \tag{4-19}$$

Which can, again using Equation 4-16 as well as:

$$\begin{aligned}\tilde{e}_{y_{food}} &= \begin{bmatrix}y_{food} \\ \dot{y}_{food}\end{bmatrix} - \begin{bmatrix}\mu_{xfood} \\ \dot{\mu}_{x_{food}}\end{bmatrix} \\ \tilde{e}_{y_{temp}} &= \begin{bmatrix}y_{temp} \\ \dot{y}_{temp}\end{bmatrix} - \begin{bmatrix}\mu_{xtemp} \\ \dot{\mu}_{x_{temp}}\end{bmatrix}\end{aligned} \tag{4-20}$$

be simplified to:

$$\begin{aligned}\frac{\partial F}{\partial \tilde{\mu}_{xfood}} &= \begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}\tilde{\prod}_w \tilde{e}_{x_{food}} + \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\tilde{\prod}_w \tilde{e}_{y_{food}} \\ \frac{\partial F}{\partial \tilde{\mu}_{xtemp}} &= \begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}\tilde{\prod}_w \tilde{e}_{x_{temp}} + \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\tilde{\prod}_{z_{food}} \tilde{e}_{y_{food}}\end{aligned} \tag{4-21}$$

Which can then be formulated as:

$$\begin{aligned}\dot{\mu}_{x_{food}} &= -\frac{\partial F}{\partial \mu_{x_{food}}} \\ \dot{\mu}_{x_{temp}} &= -\frac{\partial F}{\partial \mu_{x_{temp}}}\end{aligned} \tag{4-22}$$

Gradient Descent on action

Having defined the gradient descents on the state estimates. The next step that can be taken, is the determination of action. As explained before, action is what determines the sensory states and is thus what closes the loops in Hydar's model architecture. From Chapter 2, 2-21 ,and 2-22 can be used to determine the gradient descent functions on action:

$$\dot{u} = -\frac{\partial F}{\partial u} = -\frac{\partial \tilde{y}_{food}}{\partial u}^{\top} \frac{\partial F}{\partial \tilde{y}_{food}} - \frac{\partial \tilde{y}_{food}}{\partial u}^{\top} \frac{\partial F}{\partial \tilde{y}_{food}} \tag{4-23}$$

$\frac{\partial F}{\partial \tilde{y}_{food}}$ And $\frac{\partial F}{\partial \tilde{y}_{food}}$ can be written as below, again using prediction errors from Equation 4-20, but this time deriving with respect to $\tilde{y}$:

$$\begin{aligned}\frac{\partial F}{\partial \tilde{y}_{food}} &= \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\prod_{z_{food}} \tilde{e}_{y_{food}} \\ \frac{\partial F}{\partial \tilde{y}_{temp}} &= \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\prod_{z_{temp}} \tilde{e}_{y_{temp}}\end{aligned} \tag{4-24}$$

Then, taking into account the relations between $\dot{x}$ and u as specified by Equation 4-1, together with Equation 4-2, $\tilde{y}$ can be formulated as:

$$\tilde{y} = \begin{bmatrix}y\\ \dot{y}\end{bmatrix} = \begin{bmatrix}g_{gp}(x)\\ g'_{gp}(x)\cdot\dot{x}\end{bmatrix} = \begin{bmatrix}x\\ \dot{x}\end{bmatrix} = \begin{bmatrix}x\\ u\end{bmatrix} \tag{4-25}$$

Using this relation, our derivative of $\tilde{y}$ with respect to action u can be formulated as:

$$\frac{\partial \tilde{y}_{food}}{\partial u}^{\top} = \begin{bmatrix} 0 & 1 \end{bmatrix}$$
$$\frac{\partial \tilde{y}_{temp}}{\partial u}^{\top} = \begin{bmatrix} 0 & 1 \end{bmatrix} \tag{4-26}$$

Equation 4-26 can be combined with 4-24 to reformulate Equation 4-23 into:

$$\dot{u} = -\frac{\partial F}{\partial u} = -\frac{\partial \tilde{y}_{food}}{\partial u}^{\top} \frac{\partial F}{\partial \tilde{y}_{food}} - \frac{\partial \tilde{y}_{food}}{\partial u}^{\top} \frac{\partial F}{\partial \tilde{y}_{food}}$$
$$= -\begin{bmatrix} 0 & 1 \end{bmatrix} \prod\nolimits_{z_{food}} \tilde{e}_{y_{food}} - \begin{bmatrix} 0 & 1 \end{bmatrix} \prod\nolimits_{z_{temp}} \tilde{e}_{y_{temp}} \tag{4-27}$$

Using this formulation Hydar determines and updates its action input for every timestep. Important to mention is that for deriving the gradient descent on action, only the exteroceptive parts are used. The relation between $\dot{y}$ and u, through $\dot{x}$ as described in Equation 4-26, is not applicable for the interoceptive sensors, as they have only encompass 0th order relations and do not contain the time derivative relation $\dot{x}$ = u. More simply:

$$\frac{\partial \varphi_{food}}{\partial u} = 0$$
$$\frac{\partial \varphi_{temp}}{\partial u} = 0 \tag{4-28}$$

This is why, for this implementation, only the exteroceptive sensory signals are included in Equation 4-23 and are thus not taken into account in the determination of the action variable.

4-3-3 Back to Hydars model architecture

Having now covered the gradient descents for the hidden state estimates as well as action it can again be slotted in the model architecture as shown in Figure4-8. Taking into account what has been discussed in Section 4-2-2, and looking especially at the blocks containing the derivatives for the gradient descents, it becomes more clear how the estimates of the states are determined and what is meant by the "weighted average" of different prediction errors as has been discussed in that section. Marked with different colours, the figure shows which prediction errors are used to determine which variables (μ_{food}, μ_{temp}, $\mu_{x_{food}}$ and $\mu_{x_{temp}}$). Inspecting the gradient descent derivative equations, one can see that μ_{food} and μ_{temp} are determined using the ε's and the e_x's of their respective modality (food or Temp) and will thus be evaluated somewhere between the priors μ_P and the sensory signals φ (and indirectly and thus less prominently y).

The idea behind this, construction is that μ is going either going to be pulled more towards its prior μ_P or towards its respective interoceptive sensory signal φ. As will be clarified in the next section this will be determined by their associated precisions. The crux here is that this interoceptive state estimation of μ can be viewed in two ways. First of all, it can be viewed as an estimate of the interoceptive state predominantly using φ. Another interpretation can in line with the Fristonian idea in active inference that predictions are used to drive action [18]. Here instead of μ being an estimate of the interoceptive states, it is used as a state where Hydar wants to be or move towards, rather than where it thinks it is. The precisions of the priors μ_P versus the precisions of the interoceptive sensory states φ is what determines which of these two ways is emphasised. More simply put, the stronger the precision on the priors, the stronger the urge to fulfil those priors and the more (by contextualising interoceptive beliefs to the external, which initiates action) Hydar will move towards this prior.

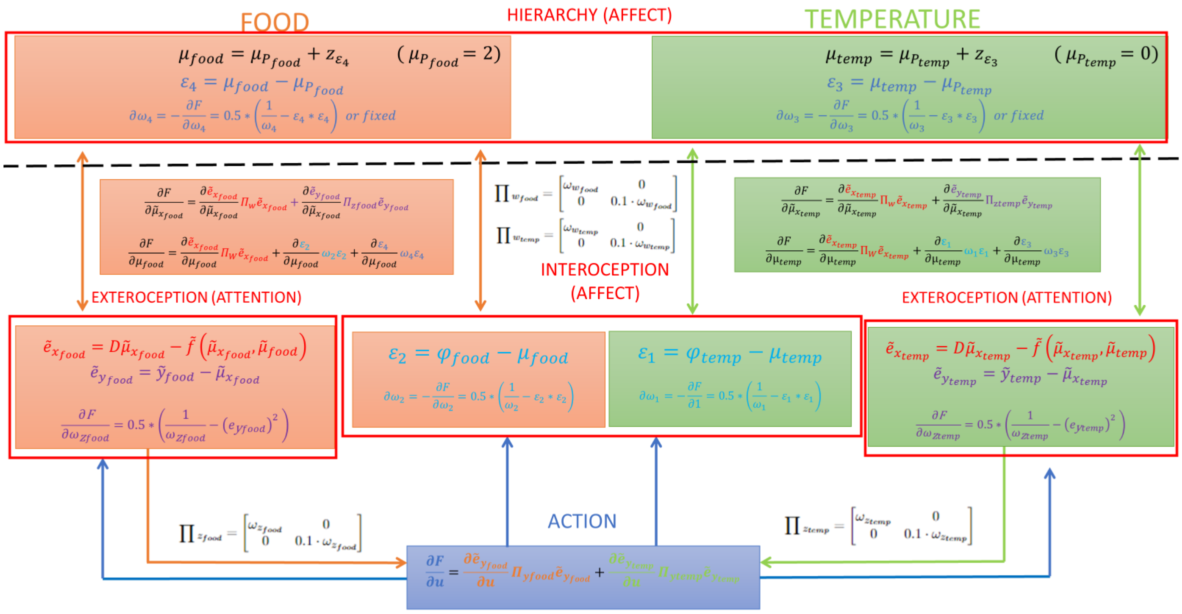


Figure 4-8: Here the gradient descent equations are slotted in at the appropriate place in the model. The colours are used to help clarify which prediction errors are used in what gradient descent equations.

Crucial in this model, is that there are two modalities, food and temperature, which will have a different prior preference (0 vs 2). Hydar is going to try to make μ_{food} move towards $\mu_{P_{food}}$ and μ_{temp} towards $\mu_{P_{temp}}$ and consequently try to make the sensors match through action. However, as both priors have their optimum at a different depth, this poses an unsolvable discrepancy. It cannot be at both a depth of 0m and 2m at the same time. The prioritization of these two modalities is done by setting the values for the precisions. This is the final part of the model as well as what connects it to Solms' story. It will be discussed in the next chapter.

4-4 Prior precisions and Sensory precision estimation

4-4-1 Introduction

As described in Chapter 4, Solms associates affect with the estimation of precision ω, by taking a gradient descent of free energy with respect to this ω. This has been described by Equation 3-1c. As clarified in Chapter 2, precision acts as a gain control over the prediction errors. Chapter 3 points out that Equation 3-1c determines this precision by comparing its inverse (the variance) to the incoming prediction errors. Thus using this equation, an agent adjusts precision such that it conforms to the magnitude of the incoming prediction errors. Consequently, using Equation 3-1c, an agent can determine the gain control over a prediction error, according to the expected precision of the corresponding value, i.e. how high it expects the prediction error to be.

As highlighted in Chapter 3, this precision updating scheme is the third mechanism used to minimize free energy, next to 1. adjusting predictions such that they match sensory signals and 2. optimizing action such that the sensory signals match predictions.

For Hydar the addition of this precision optimization means that it can estimate the precision of prediction errors and prioritise them accordingly. Section 4-4-3 will elaborate on the specifics. However, firstly the precision estimations need to be defined.

4-4-2 Precision estimation equations

Hydar has two sets of sensors. One exteroceptive set and one interception set. Hydar is equipped with sensory precision estimations for both sensors:

Interoceptive The interoceptive sensors are of order p=0, meaning that they only contain the sensory information that is associated with position x. Using Equation 3-1c, the precision estimation function can be written as:

$$\begin{aligned}\dot{\omega}_1 &= -\frac{\partial F}{\partial \omega_1} = \frac{1}{2} \cdot \left(\omega_1^{-1} - \varepsilon_1 \cdot \varepsilon_1\right) \\ \dot{\omega}_2 &= -\frac{\partial F}{\partial \omega_2} = \frac{1}{2} \cdot \left(\omega_2^{-1} - \varepsilon_2 \cdot \varepsilon_2\right)\end{aligned} \tag{4-29}$$

As mentioned in Section 4-2-1, no noise is added to the interoceptive sensors, which means that the estimate precisions ω_1 and ω_2 purely consist of the discrepancies between the interoceptive sensors φ and the internal state estimates μ.

Exteroceptive The exteroceptive sensors also have a 1st order term:

$$\begin{aligned} \tilde{e}_{y_{food}} &= \begin{bmatrix} e_{y_{food}} \\ \dot{e}_{y_{food}} \end{bmatrix} = \begin{bmatrix} y_{food} - \mu_{food} \\ \dot{y}_{food} - \dot{\mu}_{food} \end{bmatrix} \\ \tilde{e}_{y_{temp}} &= \begin{bmatrix} e_{y_{temp}} \\ \dot{e}_{y_{temp}} \end{bmatrix} = \begin{bmatrix} y_{temp} - \mu_{temp} \\ \dot{y}_{temp} - \dot{\mu}_{temp} \end{bmatrix} \end{aligned} \tag{4-30}$$

This means that $\tilde{e}_y$ needs to be accompanied by a 2x2 precision matrix $\prod_z$. Active inferencein its fullest form, allows for such a matrix to be constructed, by using the variance of the 0th-order variable. The higher-order entries of the precision matrix are then determined by multiplying the variance with a constant that is determined by the smoothness parameter of the noise. A full explanation of this process including the derivation of a correct precision matrix can be found in [19].

This research does not focus on noise smoothness. Therefore the assumption has been made that the constant for the first-order term is 0.1. This term has been chosen, to ensure that first-order information is conveyed through the model to some degree, which is necessary as the action variable is dependent only on first-order terms (Equation 4-27). This constant of 0.1 has been derived through a trial and error process and does not correspond to the smoothness of the added white noise. Looking into this would be a topic for future research. The precision estimation and corresponding matrix can thus be formulated as:

$$\begin{aligned} \frac{\partial}{\partial t}\omega_{z_{food}} &= -\frac{\partial F}{\partial \omega_{z_{food}}} = \frac{1}{2} \cdot \left(\omega_{z_{food}}^{-1} - e_{y_{food}} \cdot e_{y_{food}}\right) \\ \frac{\partial}{\partial t}\omega_{z_{temp}} &= -\frac{\partial F}{\partial \omega_{z_{temp}}} = \frac{1}{2} \cdot \left(\omega_{z_{temp}}^{-1} - e_{y_{temp}} \cdot e_{y_{temp}}\right) \end{aligned} \tag{4-31}$$

Where the precision matrices are formulated as:

$$\begin{aligned} \prod{}_{z_{food}} &= \begin{bmatrix} \omega_{z_{food}} & 0 \\ 0 & 0.1 \cdot \omega_{z_{food}} \end{bmatrix} \\ \prod{}_{z_{temp}} &= \begin{bmatrix} \omega_{z_{temp}} & 0 \\ 0 & 0.1 \cdot \omega_{z_{temp}} \end{bmatrix} \end{aligned} \tag{4-32}$$

Unlike the interoceptive sensors, the exteroceptive sensors are fed with noise (this will only be the case in the second experiment in the next chapter, as the first. The ω_z's that are estimated thus consist of both the delta between y and μ_x as well as added noise.

Priors The precisions corresponding to the prior preferences are either optimized using the formula or hard-coded depending on which of the two experiments (needs vs needs or needs vs opportunity). When hard-coded, the precision can be either encoded as a constant or a function of time

$$\dot{\omega}_3 = -\frac{\partial F}{\partial \omega_1} = \frac{1}{2} \cdot \left(\omega_3^{-1} - \varepsilon_3 \cdot \varepsilon_3\right)$$
$$\dot{\omega}_4 = -\frac{\partial F}{\partial \omega_2} = \frac{1}{2} \cdot \left(\omega_4^{-1} - \varepsilon_4 \cdot \varepsilon_4\right) \tag{4-33}$$

or

$$\omega_3 = hard - coded$$
$$\omega_4 = hard - coded \tag{4-34}$$

Hidden state Dynamics Lastly, the precision matrices associated with the dynamics of μ_x are fixed beforehand and defined like below. In theory, this could also be specified using a gradient descent. This goes beyond the purpose of this research.

$$\prod_{w_{food}} = \begin{bmatrix} \omega_{w_{food}} & 0 \\ 0 & 0.1 \cdot \omega_{w_{food}} \end{bmatrix}$$
$$\prod_{w_{temp}} = \begin{bmatrix} \omega_{w_{temp}} & 0 \\ 0 & 0.1 \cdot \omega_{w_{temp}} \end{bmatrix} \tag{4-35}$$

Where $\omega_{w_{food}}$ and $\omega_{w_{temp}}$ are hard-coded and constant.

Even though this section has been short compared to the earlier sections describing Hydar's model, the functions described here are crucial as they provide an implementation of Solms' story. The exteroceptive precision updating scheme is what Solms associates with "perceptual consciousness" and the interoceptive updates are associated with "Affective consciousness". To get a better grasp on how Hydar demonstrates these two concepts, the precisions are slotted in Hydar's model as a final addition.

4-4-3 Back to Hydar's model architecture

With the precision optimization functions defined, they can be slotted in the model architecture. Before doing this, one can take a look again at the two components that constitute affect.

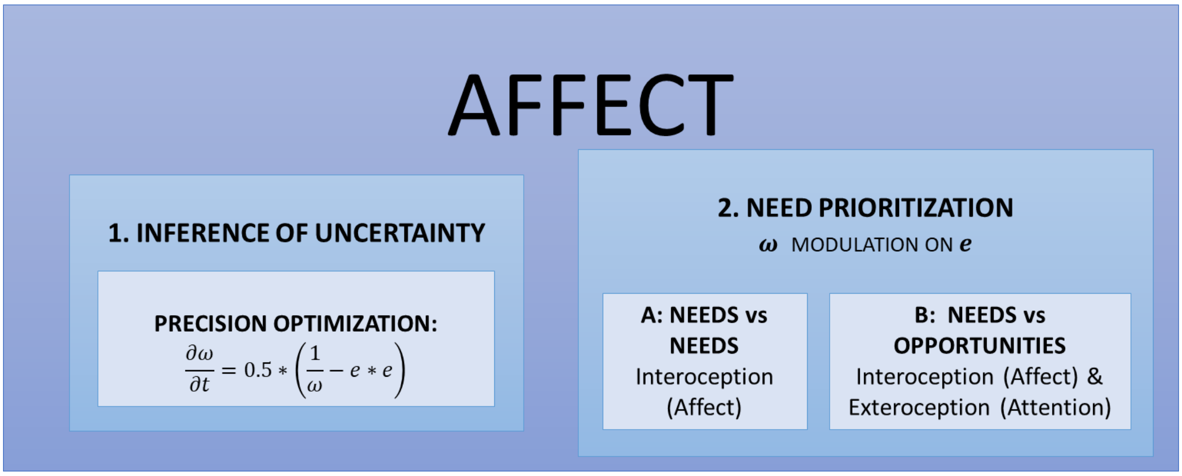


Figure 4-9: Definition of affect

As determined in Chapter 3 affect as described by Solms consists of two main parts:

1. Inference of uncertainty through precision optimization, which determines the valence of affect (positive or negative).
2. Need prioritization, as a result of precision modulation on prediction error, determining the strength of the affective signal. As highlighted this happens in two ways:
   (A) The prioritization of needs in relation to other needs.
   (B) The prioritization of needs in relation to contextual opportunities and restrictions.

Together with these definitions, the precision optimization equations can be slotted in the model architecture and affect can be highlighted in the model.

Final affective model

Figure 4-10, shows where the estimated precisions are determined and used in the model. This time, the colours indicate wich prediction errors and their associated precisions are used where. Section 4-8 ended by stating the precision estimations enable Hydar to prioritize one modality over the other. Two main mechanisms at play cause this prioritization. Firstly the prior precisions ω_4 and ω_3 have a large influence on which modality is prioritized. As explained earlier in sections 4-2-2 and 4-3-3, μ_{food} and μ_{temp} are determined through a gradient descent using their respective ε's (and $\tilde{e}_y$ through $\tilde{e}_x$). μ Will then be evaluated somewhere in between the sensory signals and the prior. Manipulation of the prior precision ω_4 and ω_3 determines how much μ_{food} and μ_{temp} are going to be pulled towards it. In other words, the prior precision determines how much Hydar is going to set its internal estimate to where it wants to be, as opposed to where it currently is according to its sensors.

Now taking Figure 4-9 in mind, there are two ways where the precisions ω_3 and ω_4 are a representation of the definition of affect. First of all the estimation of the precisions (in the case they are not hard-coded) using ε_3 and ε_4 reflect the first characteristic of affect. Here the initial values of the precisions determine whether the prediction errors are registered as positively or negatively valenced affect. The precisions also act as a gain modulator on the prediction errors, regulating their potential for arousal and thereby play a crucial role in the need prioritization. Since ε_3 and ε_4 are associated with the estimation of the interoceptive states, μ_{food} and μ_{temp} it represents "Affective consciousness". Furthermore, the specific need prioritization represented here is A: needs vs needs.

"Affective consciousness" is also represented by the interoceptive sensory signals φ_{food} and $varphi_{temp}$ and their precisions ω_1 and ω_2. Like the priors the estimation of these precisions determine valence and the gain modulation on the prediction errors ε_1 and ε_2 reflects arousal and thus need prioritisation.

The second mechanism where precision determines action, resides in the exteroceptive sensory precision updates ω_z. The crux here has everything to do with the addition of noise to the exteroceptive sensory signals y_{food} and y_{temp}. When noise is fed to one of the sensors, the precision updating formulas pick this up and decrease precision. This will then decrease the impact this sensor has on action. Hydar will be moving towards the prior in the modality where its exteroceptive sensors provide the clearest information. Taking Figure 4-9 into account, again the two aspects of affect are clearly represented in this part of the model. Precision optimization plays a big role as it is used to estimate $\omega_{z_{food}}$ and $\omega_{z_{temp}}$. These are the used in the modulation on their respective prediction errors $e_{y_{food}}$ and $e_{y_{temp}}$, which comes back to their role in need prioritization, where this time B: needs vs opportunities is represented. As this revolves around exteroception this part reflects "perceptual consciousness".

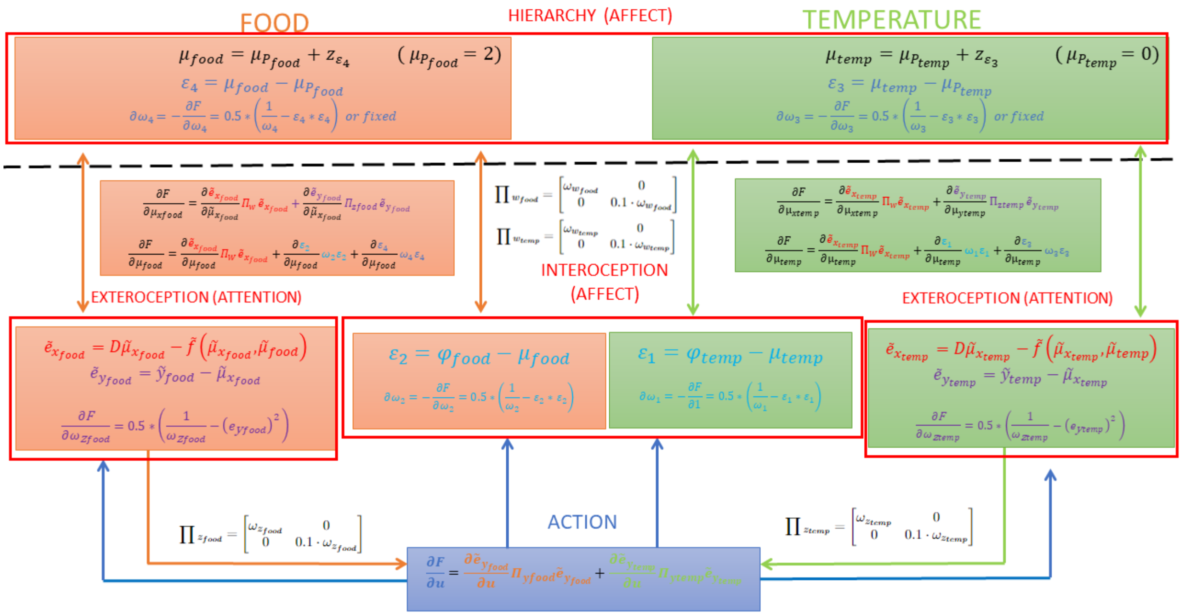


Figure 4-10: Here the precision optimization functions are slotted in at appropriate places in the model. The colours indicate which prediction errors are associated with gradient descent equations, this time including the gradient descents on precision.

Take note of how, in this implementation, action is only directly connected to the exteroceptive sensory signals. These are thus the only direct influences on action. The role of the interoceptive sensors are however: 1. To play a role in the determination of the μ's together with the priors, (and to a small extent the intero- and exteroceptive sensors), which has an indirect influence on eventual action. 2. Its second role is to report interoceptive changes in precision.

4-5 Answer to sub-question 4

This chapter answers the fourth sub-question by providing the 4 parts that the model, introduced as Hydar needs in order to correctly demonstrate affective active inference. That is:

1. Hydar needs **multiple** interoceptive needs that can compete.
2. Hydar needs an **Interoceptive** system that can measure and prioritize interoceptive prediction errors.
3. Hydar needs **Exteroception** to prioritize these needs according to context.
4. Hydar needs **action** to perform according to this prioritization.

The implementation of these 4 requirements can be found in Figure 4-3, with a more detailed account in Figures 4-7, 4-8 and 4-10. With the model in place, one can go on to the following chapter, which dives into a practical implementation of the model.

Chapter 5

Experiments and Results

5-1 Introduction

Chapter 4 shows that the four model design requirements for Hydar that were introduced in Section 4-1, are fulfilled with the final model. This chapter focuses on the model behaviour and contains 2 core experiments that are each designed to illustrate specific characteristics of the model behaviour as specified in Section 3-4. The two characteristics are:

1. Inference of uncertainty through precision optimization
2. Need prioritization, including:
   (A) The prioritization of needs in relation to other needs.
   (B) The prioritization of needs in relation to contextual opportunities and restrictions.

The results will show that Hydar can perform all of this successfully. This will be illustrated with graphs that depict Hydar's position alongside the state estimates μ_{food} & μ_{temp} and $\mu_{x_{food}}$ & $\mu_{x_{temp}}$. Furthermore, the estimated variances are depicted accompanied by the precision, which are their inverses.

Aside from presenting these results in the form of graphs accompanied by an in-depth explanation this chapter will also link these simulation results back to the established definition of affect and consciousness.

The next two paragraphs will provide an introduction to the two experiments. Followed by Section 5-2 which focuses on illustration of the results and an in-depth analysis. By the end of this chapter, it will be clear that Hydar is indeed able to perform precision optimization and prioritization through both mechanisms A and B.

1. Conflicting prior preferences

As shown in Section 3-4, needs are partly prioritized in relation to other needs and partly in relation to context. This experiment focuses on the first of the two. It is designed to illustrate the importance of the hierarchical priors. More specifically, it focuses on the effect of the precisions of the priors on Hydar's behaviour. The value of this precision, or more precisely, the relative values of the precisions, determines which modality (food or temperature) is prioritized. Or in Solms terms: which affect is going to be felt more strongly? In the case of this experiment, the food precision will increase linearly, while the temperature prior will stay constant. This could be interpreted as a representation of Hydar's increasing hunger over time. For the purpose of illustration, this experiment uses a simplification with a pre-defined increase in precision on the food prior. That is, the precisions of the priors are not determined using Equation 3-1c. Nonetheless, the result, which is discussed in the next section, will show that Hydar adapts its depth according to the afforded precision on the priors. In a more sophisticated affective model, that could represent a more realistic organism, this increasing food precision might not be hard coded, but rather a result of the precision optimisation that is driven by a much larger and more complete hierarchical structure. In contrast with the priors, the precisions of the interception (and exteroceptive) sensors are not predefined but determined with the precision update formula. This is done to monitor valence with respect to interoceptive sensors.

2. The difference in sensory reliability

This experiment illustrates precision estimation of exteroceptive signals and how the needs are prioritized with respect to opportunities and how this influences Hydar's behaviour. As highlighted in Section 3-4, Solms argues that the prioritization of needs (through affect) depends crucially on unfolding context and its opportunities. In the case of this simulation, opportunities are represented by the amount of noise that the exteroceptive sensors pick up. Hydar's goal is to minimize its free energy using the information that gives it the best ability to do so. A noisy sensor provides a less clear representation of the outside world, which leaves Hydar with a weaker capability of reducing its free energy. To account for this, Hydar needs to be able to emphasize the sensory signals that provide the clearest information.

Imagine that the temperature sensor could give a blurry reading due to turbulence in the water, with rapid temperature fluctuations. The food sensor on the other hand could give a blurry reading when visibility in the water goes down and Hydar cannot spot the bottom (where the food is located) anymore. Turbulent water and a noisy temperature reading, will result in Hydar downregulating the precision of this sensor and thus reducing its influence. In other words, Hydar will be directing his attention to his food sensor. Consequently it will move down towards the bottom at 2m to comply with its food needs. The opposite will happen if the food sensory gives a noisy reading. That is, Hydar is not able to see very well and will reduce its sensory precision and prioritize its temperature reading and thus move towards its preferred temperature depth. This is demonstrated in the second experiment and concerns the attention part of the model.

Simultaneously with attention, the experiment shows precision updates of the interoceptive sensory signals and slightly with the priors. These updates are associated with the affective part of the model. This will all be illustrated and clarified in the next section.

5-2 Results

In this section, the model results of the previously explained implementations are shown accompanied by an in-depth analysis.

5-2-1 Conflicting prior preferences

This experiment is mostly focused on showcasing the first way of need prioritization: Prioritizing needs in relation to other needs. As mentioned earlier, note that the increase of the prior precision in this simulation is hard-coded. This means that, concerning the priors, the precision optimization of Equation 3-1c is not used. For this reason, this experiment does not fully showcase the first characteristic of affect, precision optimization. Nonetheless, this function is still used in the interoceptive sensors φ_{food} and φ_{temp}. So it still showcases this formula to some extent in the results.

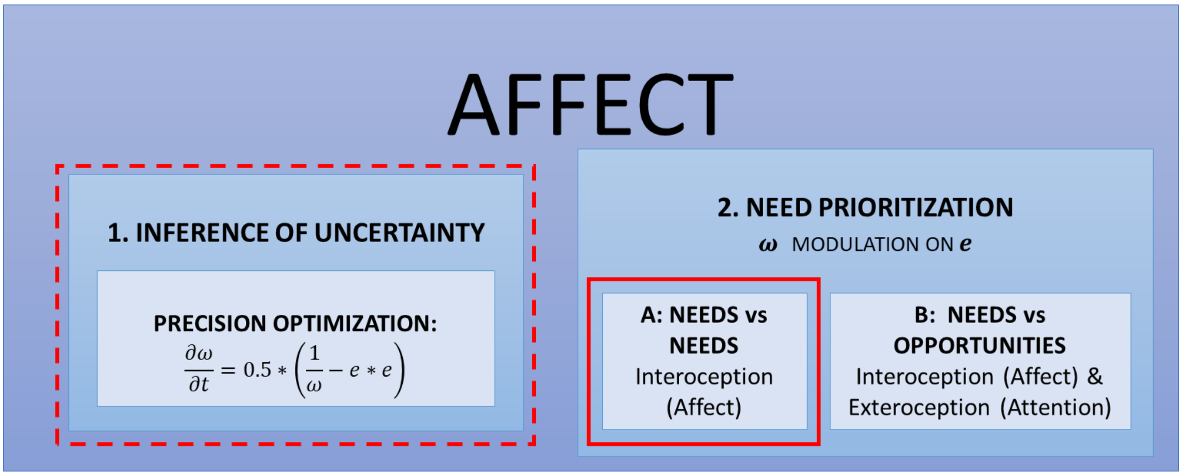


Figure 5-1: The first experiment focuses slightly on the precision optimization formula. It focuses mostly on prioritising needs in relation to other needs.

As explained in sections 4-2-2 and 4-3-3 accompanied by the figures 4-8 and 4-10, μ_{food} and μ_{temp} are determined as a weighted average between their respective sensors (interoceptive and exteroceptive) and their respective priors. Prior precision influences how tightly the μ's are pulled towards their prior. In turn, the $\mu's$ are used as a causal state in the Attractor Dynamics functions f, which ultimately determines where Hydars action input wants to steer towards. Action uses exteroceptive prediction error for this.

This means that the precisions on the priors (ω_4 on $\mu_{P_{food}}$ and ω_3 on $\mu_{P_{temp}}$) dictate how strongly their respective μ's are pulled towards their preferred value. Thus, a prior with a stronger precision has a stronger pull on its respective μ, allowing for less deviation. As the μ's are used as causal states (or attractor states), the prior precision has an influence on action. In this experiment, this means that the increasing precision of $\mu_{P_{food}}$ results in a stronger pull on μ_{food}, resulting in decreasing allowance for discrepancy between $\mu_{P_{food}}$ and μ_{food}.

The simulation results, as shown in Figure 5-2, clearly confirm the behaviour just characterised. As a reminder, in this simulation (as well as the next experiment) the ideal food value is 2 and the ideal temperature value is 0 (note that the value of the variable is equal to its depth so ideal depths are 0m and 2m respectively).

To keep it simple, no noise is added to this simulation. The precision of the temperature prior is set at a value of 3 throughout the whole simulation. The food precision starts at 0.3 (factor 10 lower), from t=250, the precision starts rising linearly up until 30 (factor 10 higher). After t = 750, the precision is instantly decreased to 3 (equal to Temp precision), which is done for illustrative purposes of the model behaviour. Nevertheless, it could symbolise Hydar having obtained enough food and thus deciding that Hunger is not to be prioritised anymore.

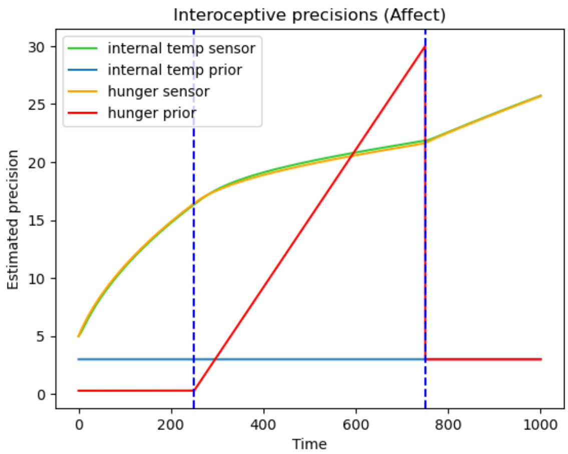
	food prior	**Temp prior**	**Internal food sensor**	**Internal Temp sensor**	**External food sensor**	**External Temp sensor**
Variable symbol	$\mu_{P_{food}}$	$\mu_{P_{temp}}$	φ_{food}	φ_{temp}	y_{food}	y_{temp}
Associated precision	ω_4	ω_3	ω_2	ω_1	$\omega_{z_{food}}$	$\omega_{z_{temp}}$
(starting) value	1. 0.3-30 2. 3	3	5	5	5	5
$\frac{\partial\omega}{\partial t}$ **/ hard-coded**	hard-coded	hard-coded	$\frac{\partial\omega}{\partial t}$	$\frac{\partial\omega}{\partial t}$	$\frac{\partial\omega}{\partial t}$	$\frac{\partial\omega}{\partial t}$

Table 5-1: Precision values for experiment 1

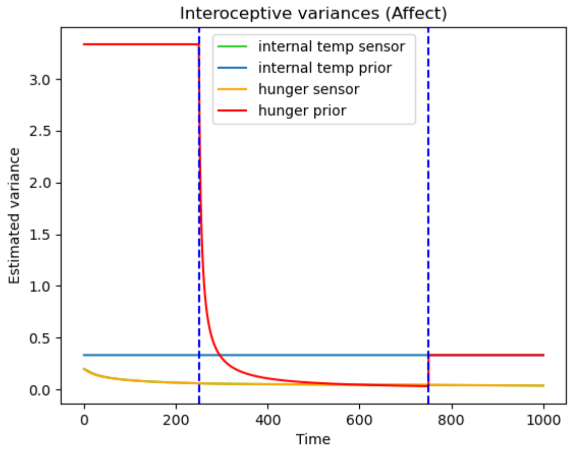
The increasing food precision is illustrated by the red line in figures: 5-2a and 5-2b. Figure 5-2c, shows Hydar's position over time (red line) and the external estimates $\mu_{x_{food}}$ and $\mu_{x_{temp}}$ which drive action.

As expected, Figure 5-2c shows that initially, when the precision on the food prior is low, Hydar moves towards its preferred temperature prior, indicating that prior precision determines the depth that Hydar wants to and will go to. After t = 250, the precision on the food prior starts rising. As a result, Hydar starts moving downwards, closer to its food prior. Due to the food prior precision exceeding the precision of the temperature prior, Hydar moves downwards beyond the middle (1m), which again demonstrates that indeed a larger prior precision influences the behaviour. Then when the priors are set equal, Hydar moves towards an equilibrium that is right in the middle at a depth of 1m.

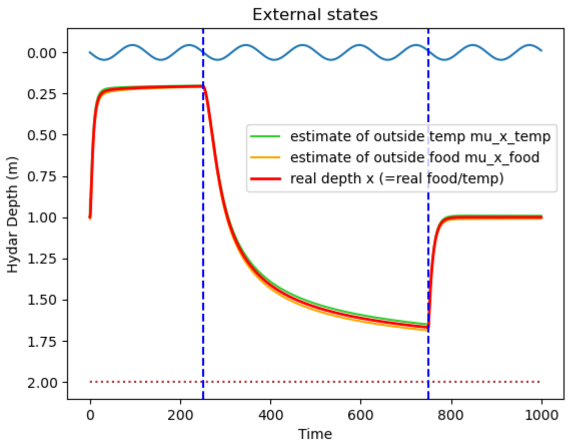
Figure 5-2d, shows the internal measurements that are a result of Hydar's depth, in red and blue, as well as the internal estimates μ_{food} in orange and μ_{temp} in green. This figure particularly shows the importance of the prior precision. In the first section (until t=250), the temperature prior has a higher precision than the food prior, meaning that μ_{temp} is pulled towards $\mu_{P_{temp}}$ more strongly than μ_{food} is pulled towards $\mu_{P_{food}}$. That is, the green line is pulled towards its preferred depth of 0m, whereas the orange line is not tightly bound to its prior, and thus allowed to move away from its preferred depth. This, through the exteroceptive parts of the model, results in Hydar's movement as seen in the graph. After t=250, the food prior precision increases, making it less and less forgiving, resulting in the orange line being pulled more towards a depth of 2m.



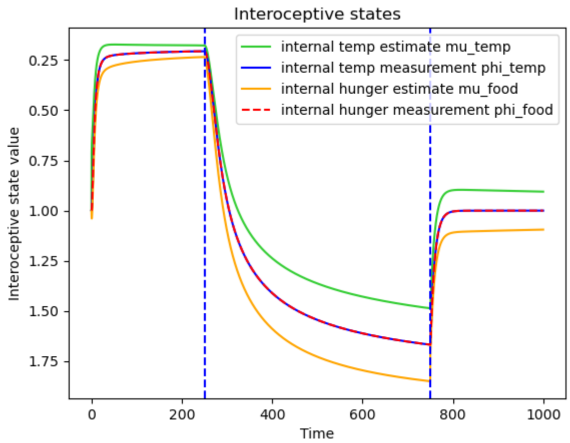
(a) The red line shows the hard-coded increasing hunger prior precision ω_4. The blue line shows the temperature prior precision ω_3, which remains constant.



(b) This shows the inverses of the variables as Figure a. In other words, it shows the interoceptive variances.



(c) The red line shows the position x of Hydar in the water as a function of time. The green and orange line represent the exteroceptive state estimates $\mu_{x food}$ and $\mu_{x_{temp}}$



(d) This figure shows the interoceptive sensory states φ and the internal state estimates μ

Figure 5-2: The results of the first experiment. Here (a) shows the precisions over time, (b) shows the variances over time, (c) shows the position x over time as well as the exteroceptive state estimates and (d) shows the interoceptive measurements and estimations.

Note in a more fully worked out representation, with a deeper hierarchy, the prior precisions could possibly be optimised using Equation 3-1c. However, changing the precisions manually shows proof of behaviour that is very well in accordance with Solms' definition of the functions of affect. That is, "prioritising needs in relation to other needs". The experiment shows very clearly that differences in prior precisions allow Hydar to prioritize one type of behaviour over another. From a line of reasoning more in line with the "Affective" story of this research, the increase of the precision of $\mu_{P_{food}}$ influences how the discrepancy (or prediction error) between μ_{food} and $\mu_{P_{food}}$ is perceived. That is, the higher the precision on the prior, the more salient prediction errors on its respective variable are and the lower the prediction error needs to be for it to be registered as negatively valenced affect. Thus, the increase of precision, makes the prediction error ε_4 come increasingly forwards as arousal, ensuring Hydars need to comply to its food preference of being at a depth of 2m arises to its affective awareness.

It could perhaps be conceivable that in a more complete affective active inference model, with a deeper hierarchy, these prior precisions could be manipulated such that the same rise in Hunger-precision over time is achieved using Equation 3-1c at every level of the hierarchy. Another interpretation could be that organisms have the ability to infer precisions, but also to direct them according to necessity. More of this is discussed in the next chapter as this could prove an interesting topic for future research.

Note that even though it is not the central point made in this experiment. The precisions of the interoceptive sensors are estimated in this experiment, which relates back to the first aspect of affect in the Figure 5-1. The rise in this interoceptive precision indicates that Hydar is able to increasingly match it's predictions μ to their respective sensory signals φ, producing positive valence.

5-2-2 Difference in sensory reliability

This experiment is about demonstrating a working precision estimation mechanism, as well as demonstrating the second way in which interoceptive needs are prioritized: The prioritization of needs in relation to opportunities and restrictions posed by the environment.

"Opportunities and restrictions" in this simulation are represented by noise in the sensory signals. Here a signal with less noise represents more opportunity and vice versa. The key concept at play in this experiment is that Hydar can properly make an estimate of the external sensor variance, adjust its exteroceptive sensor precision and use this information to both correctly adjust its interoceptive precisions (affect) and move through the environment correctly.

Thus, this experiment highlights two aspects within the definition of affect (Figure) 5-3. First of all it highlights whether Hydar can correctly infer exteroceptive uncertainty in the form of noise. Using this process, Hydar demonstrates the second characteristic by prioritizing its interoceptive needs accordingly, using exteroceptive information.

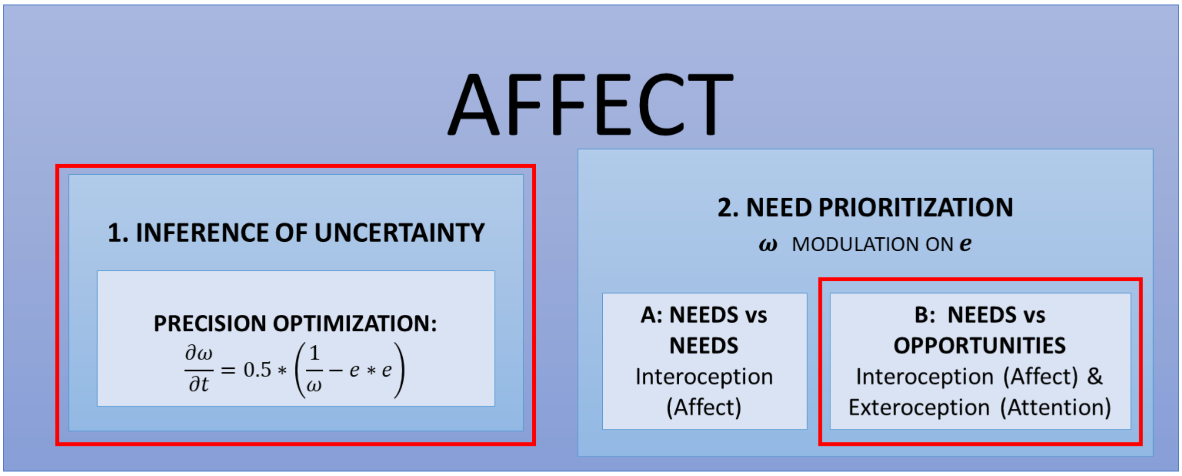


Figure 5-3: The focus of the second experiment is very much on the precision optimization formula and the estimation of exteroceptive noise. Furthermore, it highlights the prioritization of needs with respect to opportunities

To test this, Hydar, again, has two conflicting prior preferences. Its ideal food measurement is set at 2, whereas its ideal temperature measurement is set at 0 (note again that the value of the variable is equal to its depth so ideal depths are 0m and 2m respectively). This time, the priors as well as the internal and external signals are all equipped with an optimizable precision. The starting values for the precisions can be found below corresponding to Figure 4-10:

	food prior	**Temp prior**	**Internal food sensor**	**Internal Temp sensor**	**External food sensor**	**External Temp sensor**
Variable symbol	$\mu_{P_{food}}$	$\mu_{P_{temp}}$	φ_{food}	φ_{temp}	y_{food}	y_{temp}
Associated precision	ω_4	ω_3	ω_2	ω_1	$\omega_{z_{food}}$	$\omega_{z_{temp}}$
Starting value	10	10	5	5	5	5
$\frac{\partial\omega}{\partial t}$ **/ hard-coded**	$\frac{\partial\omega}{\partial t}$	$\frac{\partial\omega}{\partial t}$	$\frac{\partial\omega}{\partial t}$	$\frac{\partial\omega}{\partial t}$	$\frac{\partial\omega}{\partial t}$	$\frac{\partial\omega}{\partial t}$

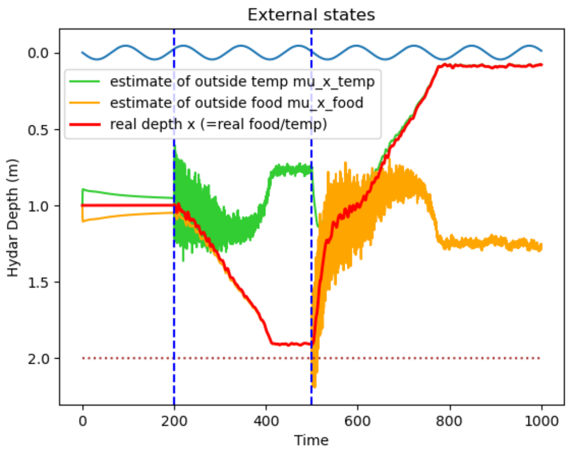
Table 5-2: Starting values of the precisions of experiment 2

As can be seen, the prior start with a higher value to make sure that Hydar's beliefs μ_{food} and μ_{temp} are indeed pulled towards their preferred values. That is, the μ's are mostly determined by comparing them to the measured interoceptive values φ_{food} and φ_{temp} and the prior values $\mu_{P_{food}}$ and $\mu_{P_{temp}}$ (i.e. the resulting prediction errors are used in the state updates). As the priors have a high initial precision, the μ's will be pulled towards the prior values more strongly. In other words, by giving the priors a high expected precision, it allows for less deviation i.e. the threshold for deviations to be registered as negative affect is lower. This is important, as μ_{food} and μ_{temp} are the attractor states for the exteroceptive dynamic updates, which will be shown in the next paragraph which will cover "perceptual consciousness" or "attention. This will then be followed by a deeper dive into "Affective consciousness"

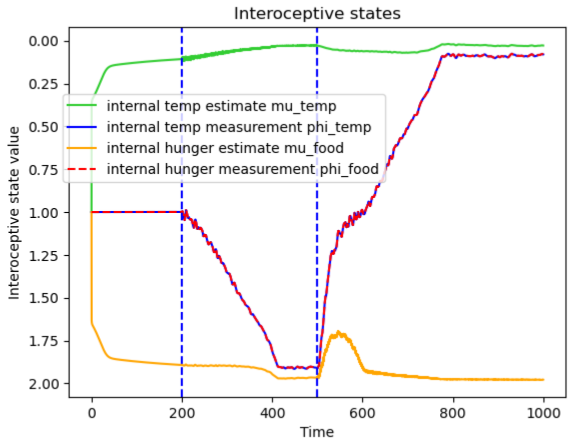
The simulation results are illustrated in Figure 5-4. The experiment has three stages. In the first stage, Hydar will start at an equilibrium depth of 1m. At t = 200, Hydar's external temperature sensor is fed noise with a variance of 0.25. This results in the two proceedings that are explained in the introduction of this experiment. The first one being: The exteroceptive precision estimation and resulting action input, which is associated with "perceptual consciousness" or attention.

Exteroceptive precision optimization/attention

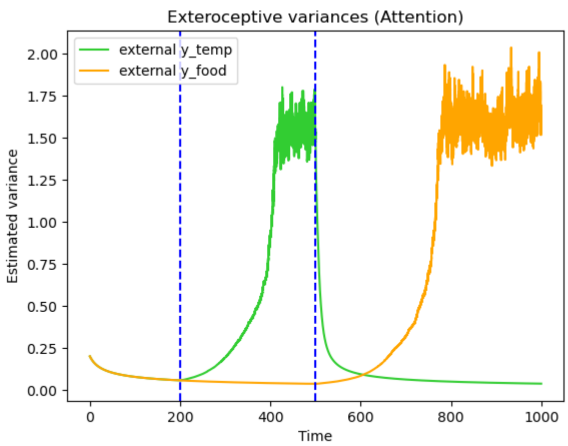
Due to estimation of the added noise, Hydar updates the precision of the exteroceptive temperature sensor (figures 5-4c and 5-4e), this results in a relatively stronger impact of the food sensor on action, which makes Hydar tend to move towards his preferred food depth (2m). This can be observed by inspecting the red line in Figure 5-4a. Without the added noise, more specifically the estimation of it, Hydar would not have made this deviation from the middle. However, as the temperature sensor now produces noisy (and thus more incoherent) information, Hydar chooses to prioritise the clearer information provided by its food sensor and moves downwards. This could represent Hydar experiencing temperature fluctuations due to turbulent water. As this restricts Hydar from staying at the right temperature level, Hydar diverts its attention towards its food sensor and moves downwards. Taking in mind Figure 5-3, it can be easily deduced that both precision optimization and need prioritization are represented here. Hydar experiences negative affect due to the increasing uncertainty



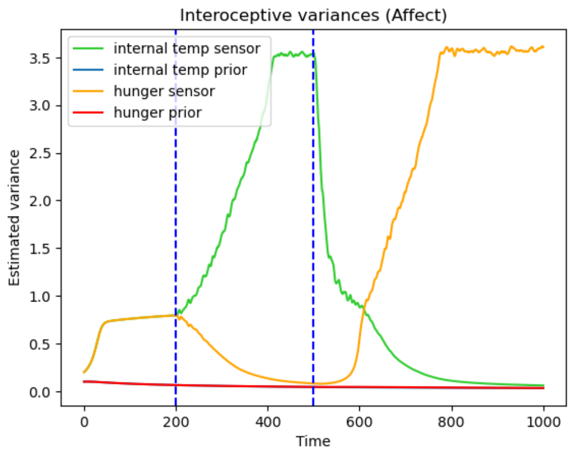
(a) This image shows the actual position of Hydar (red), accompanied by its external temperature (green) and food (red) estimations $\mu_{x_{temp}}$ and $\mu_{x_{food}}$, note how its position changes when adding noise to one of the two sensors.



(b) This image shows Hydar's internal measurements. This corresponds with its real position. The orange and green lines denote the estimates of the internal states μ_{food} and μ_{temp}

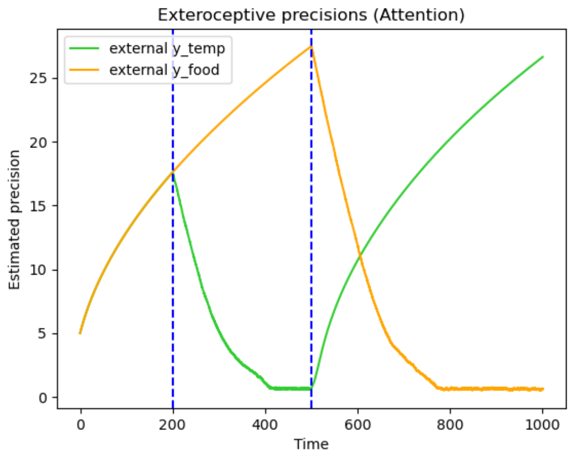


(c) This image shows the external variance estimation. Note that the estimation itself is also noisy, as it uses a noisy prediction error as input, a high variance means a low precision and thus down-modulation of that sensor (attention)

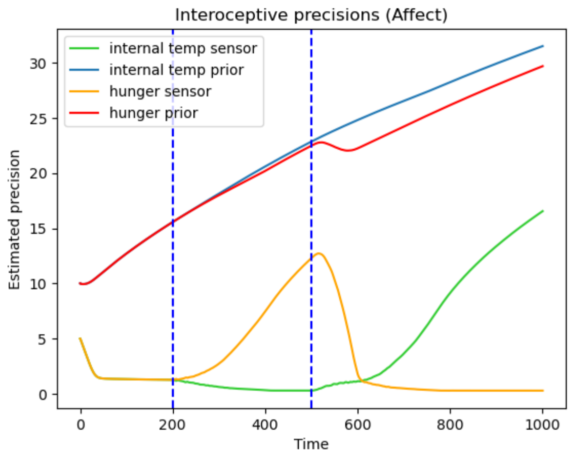


(d) this image shows the variances associated with the interoceptive sensors. A high variance means a low precision and thus down-modulation of that sensor (affect)

Figure 5-4



(e) This figure shows the external sensor precisions, which is the inverse of figure c



(f) This figure provides the precisions which is the inverse of figure d

Figure 5-4: These figures show the results of the second experiment. It demonstrates Hydar's behaviour when noise is added to each of the exteroceptive sensors one after the other. The blue lines indicate when the injection of noise starts and when it switches to the other sensor

induced by the noise. As a result, it reduces the salience of the sensor and focuses in prioritizing it's food related needs, where it is able to increase precision and thus experience positive affect in the category that is now most salient.

It should be noted that movement towards the food prior and away from the temperature prior in turn increases the sensory prediction errors (intero- & exteroceptive), of the temperature module. Additionally to the depth x, Figure 5-4a shows the estimations of the external states $\mu_{x_{food}}$ and $\mu_{x_{food}}$, depicted by the green and orange lines.

As in Hydars world, the sensory values are equal to its actual depth plus the noise, the discrepancies between those two lines and the red line together with the added noise from the exteroceptive prediction errors as described in Equations 4-5 and 4-9 which results in:

$$
\begin{aligned}
\tilde{y}_{temp} &= \tilde{g}_{gm}(\tilde{\mu}_{x_{temp}}) + \tilde{z} \\
\tilde{y}_{food} &= \tilde{g}_{gm}(\tilde{\mu}_{x_{food}}) + \tilde{z} \\
\tilde{e}_{y_{temp}} &= \tilde{y}_{temp} - \tilde{\mu}_{x_{temp}} \\
\tilde{e}_{y_{food}} &= \tilde{y}_{food} - \tilde{\mu}_{x_{food}}
\end{aligned}
$$

These discrepancies are then again represented in Figure 5-4c, where the green graph goes up as it accounts for the combination of the square of the prediction error $e_{y_{temp}}$ and the variance of the added noise. Figure 5-4e shows the inverse of this graph, the precisions, where it can be seen that the green line drops as precision decreases. The inference of this precision/variance shows negative affect in the temperature category. From this figure, it also becomes more clear that the precision of the external food estimation increases, due to Hydar adhering to its food prior. This represents positively valenced affect. Here food is now the most salient category of need with the highest arousal.

The estimations of the variance in Figure 5-4c comprise both the difference between the real depth x and the external estimate μ_x in addition to the injected noise of variance 0.25. To validate, the sum of the mean and variances of the prediction errors $e_{y_{food}}$ and $e_{y_{temp}}$ from t = 430 to t = 500 (the steady state plateau in Figure 5-4), are compared to their respective inverse precision ω_z^{-1}.

Measured variance and prediction error:	**Value**	**Average estimated inverse precision:**	**Value**
$var(e_{y_{temp}}[430:500]) + mean(e_{y_{temp}}[430:500])$	1.5419	$mean(\omega_{z_{temp}}[430:500])^{-1}$	1.5389
$var(e_{y_{food}}[430:500]) + mean(e_{y_{food}}[430:500])$	5.206e-6	$mean(\omega_{z_{food}}[430:500])^{-1}$	0.0381

Table 5-3: Result validation of exteroceptive precisions. Here it shows that the sum of prediction error and the variance of the added noises are closely estimated by their respective ω_z

As can be seen from the table. The estimated inverse precisions (or variance) indeed correspond to the sum of the prediction errors and added variances. Note that the small difference in the second row is due to the precision estimation of ω_{zfood} converging towards but not reaching 0.

From t = 500, the noise is switched from the sensory signal y_{temp} to y_{food}. This could represent visibility in the water decreasing, with the result that Hydar is not able to locate where it should be to gather food. Looking at the aforementioned figures, it can be seen that the same process happens again, this time re-evaluating the exteroceptive precisions and thus upregulating the temperature sensory signals and downregulating the food sensory signals. So now, precision in food is decreased (negative affect) and precision in temperature is increased (positive affect). Temperature is now becoming the main source of arousal. As a result, Hydar moves upwards, towards its preferred temperature value. The variances (and associated precisions) are also updated accordingly as can be seen in the aforementioned figures as well as the table below.

Measured variance and prediction error:	**Value**	**Average estimated inverse precision:**	**Value**
$var(e_{y_{temp}}[800:1000]) + mean(e_{y_{temp}}[800:1000])$	6.222e-6	$mean(\omega_{z_{temp}}[800:1000])^{-1}$	0.04266
$var(e_{y_{food}}[800:1000]) + mean(e_{y_{food}}[800:1000])$	1.5991	$mean(\omega_{z_{food}}[800:1000])^{-1}$	1.5990

Table 5-4: Result validation of the exteroceptive precisions. Here it shows that the sum of prediction error and variance of the added noise are closely estimated by their respective ω_z

Again the estimated inverse precisions correspond to the measured variances and prediction errors.

The process that has just been described, where Hydar infers exteroceptive precision and modulates the "arousal" coming forth from two different sensory signals represents "perceptual consciousness" or attention. Here it is clear that regarding Figure 5-3, the inference of precision is demonstrated, representing the valence induced by the two sensory signals. Simultaneously, this precision optimization regulates the arousal of exteroceptive prediction errors $e_{y_{food}}$ and $e_{y_{temp}}$, inducing the need prioritization where the two interoceptive needs are prioritized using exteroceptive contextual information.

This experiment also highlights the interoceptive precision estimation, which Solms associates with "Affective consciousness" or Feeling.

Interoceptive precision optimization/Affect

The results associated with the "Affective consciousness" part of the experiment are depicted in figures 5-4b, 5-4d and 5-4f. These represent the interoceptive, states, variances and precisions respectively. Interoception is performed by estimating the states μ_{food} and μ_{temp}. As explained in Chapter 4 and mentioned above this is mostly done by comparing them to the measured interoceptive sensory values φ_{food} and φ_{temp} and the prior values $\mu_{P_{food}}$ and $\mu_{P_{temp}}$. Again, as the priors have a high initial precision, the μ's will be pulled towards the prior values more strongly. In other words, by giving the priors a high expected precision, it allows for less deviation i.e. the arousal of ε_3 and ε_4 is higher and the threshold for deviations to be registered as negative affect is lower as compared to ε_1 and ε_2.

Figure 5-4b, shows the internal sensory states, note that these signals are identical to the position x in Figure 5-4a due to the relation provided by Equation 4-3, which denotes that in Hydars world, the interoceptive signals are equal to the depth. Furthermore, this figure shows the internal state estimates μ_{food} and μ_{temp} in orange and green. Similarly, as in the exteroceptive sub-modules, the discrepancies between the sensory lines and the state estimate lines illustrate prediction errors as in Equation 4-10:

$$\varepsilon_1 = \varphi_{temp} - \mu_{temp}$$
$$\varepsilon_2 = \varphi_{food} - \mu_{food}$$

Furthermore the distance from μ_{food} and μ_{temp} and their respective priors $\mu_{P_{food}}$ and $\mu_{P_{temp}}$ (2m and 0m) can also be interpreted from this figure and contributes to the prediction errors that are defined in Equation 4-11:

$$\varepsilon_3 = \mu_{temp} - \mu_{P_{temp}}$$
$$\varepsilon_4 = \mu_{food} - \mu_{P_{food}}$$

Note that these discrepancies are small as a result of the high afforded initial values for ω_3 and ω_4.

In Figure 5-4e, the internal variances are shown. Take note, that as there is no noise added to the interoceptive sensors, this estimated variance only comprises the square of the interoceptive prediction errors just described. (more precisely there will be a little bit of noise that propagated from the exteroceptive sensors to the real depth x and thus also through the interoceptive sensors.)

Measured variance and prediction error:	**Value**	**Average estimated inverse precision:**	**Value**
$var(\varepsilon_1[430 : 500]) + mean(\varepsilon_1[430 : 500])$	3.53275	$mean(\omega_1[430 : 500])^{-1}$	3.53259
$var(\varepsilon_2[430 : 500]) + mean(\varepsilon_2[430 : 500])$	0.0036	$mean(\omega_2[430 : 500])^{-1}$	0.09751

Table 5-5: Result validation of the interoceptive precisions. Here it shows that the sum of prediction error and variance of the noise that propagates through the model are closely estimated by their respective ω

Measured variance and prediction error:	**Value**	**Average estimated inverse precision:**	**Value**
$var(\varepsilon_1[800 : 1000]) + mean(\varepsilon_1[800 : 1000])$	0.003515	$mean(\omega_1[800 : 1000])^{-1}$	0.07725
$var(\varepsilon_2[800 : 1000]) + mean(\varepsilon_2[800 : 1000])$	3.575360	$mean(\omega_2[800 : 1000])^{-1}$	3.575294

Table 5-6: Result validation of the interoceptive precisions. Here it shows that the sum of prediction error and variance of the noise that propagates through the model are closely estimated by their respective ω

Again the tables show that the precision estimations match the measured variances and prediction errors.

Section 3-2 noted that every hierarchical abstraction contains an optimizable precision. Affect as registered by the agent is constituted by unexpected uncertainty in the category with the highest afforded precision within this hierarchy. From that point of view, it would be reasonable to assign all the precision optimizations in the hierarchy that are associated with interoception of ω_1, ω_2, ω_3 and ω_4 to "Affective consciousness". That is, deviations in the form of prediction errors using prediction errors ε_1, ε_2, ε_3 and ε_4 all have the potential to be felt as affect (or arousal) by Hydar. The prioritized and thus felt deviation depends on the balance of the afforded precisions. In the case of this simulation, ω_3 and ω_4 are afforded a high initial precision. Therefore, deviations from the priors, are given large weight and are thus used in the optimization and will be corrected. In other words, they come to Hydar's affective awareness and through exteroception Hydar solves this by moving towards the preferred prior. So in short, when the precisions are high at first, they tend to stay high as there is more weight given to deviations to the concerning variable. This can be seen by looking at the blue and red lines in figures 5-4d and 5-4f.

The interoceptive sensory signals φ have a lower initial precision, meaning that they are afforded less weight in the optimization. This lower initial precision results in a looser tolerance towards deviations. Concomitantly, the lower precisions give a smaller weight to their prediction errors, resulting in less influence in the optimization of μ (see Equation 4-15), which in turn results in that deviations are more likely. This then results in the updating of the green and orange lines in figures 5-4d and 5-4f. Note that the exteroceptive precisions estimations have a large influence on this. Before any noise is added, the internal states μ_{food} and μ_{temp} cause a constant conflict as they both tend to move towards their own prior, which in turn, through the exteroceptive module, makes Hydar want to move to both 0m and 2m depth. As both priors have the same precision to start with as well, this results in Hydar staying at a depth of 1m. This in turn results in that both modalities (food and temperature), will provide Hydar with an unsolvable interoceptive prediction error. To accommodate for this, Hydar increases its expected variance (and thus lowers its expected precision), which can be seen in Figure 5-4e, where the internal variances match the square of the prediction errors that can be deduced by looking at Figure 5-4b (i.e. the discrepancy between the blue and the green line and between the orange and the red line).

When at t = 200 the noise is added to the exteroceptive food sensor, this balance is disturbed and as Hydar starts moving, resulting in ε_1 to increase and ε_2 to decrease. This movement is dictated by exteroceptive opportunity (noise). Thus Hydar updates its interoceptive precisions. In other words, Hydar reduces the precision ω_1 of the sensor φ_1. This decrease in precision would result in negatively valenced affect. However, due to the external conditions, Hydar has "decided" that the category where uncertainty cannot prevail is food. This can also be seen in the figures as the variance of the food sensor decreases and the precision increases. So in short, Hydar opted for an increase in precision and thus positive affect in the most salient category which happens to be food. After t=500, this whole process is reversed as can be seen in the figures.

5-3 Conclusion

The main goal of this chapter was to show in a simulation that using the precision optimization formula from equation $3-1c$, it is possible to implement a model that can perform the two tasks that were presented in Section 3-4:

1. Inference of uncertainty through precision optimization
2. Need prioritization, including:
   (A) The prioritization of needs in relation to other needs.
   (B) The prioritization of needs in relation to contextual opportunities and restrictions.

The simulation presented in this chapter has successfully implemented this and thus proven that it is indeed possible to perform these two tasks. It shows how top-down modulation of the precisions can induce certain actions. Furthermore, this model gives some insight as to what the implementation of both "Affective " as well as "perceptual" consciousness could look like in an active inference model. It is very important to realise that this is a heavily simplified model and should not be considered affective to the extent real life organisms are. For the same reasons it can also not be called conscious. Nonetheless, this implementation and its outcomes give some insight into how the mechanisms that according to Solms are the foundation of affect and consciousness work. Furthermore, it could be used in more sophisticated models in an effort to create an agent that could potentially simulate real affective processes.

On account of adaptability, this model proves that the optimization of precisions and thus affect can indeed provide an agent with advantages. First of all the tuning of the prior precisions allows the agent to prioritize its homeostatic needs and act upon that prioritization. Furthermore, the simulation shows adaptable behaviour towards exteroceptive noise to some degree. Here Hydar moves towards the prior of the sensory modality that receives the least amount of noise. The results thus provide an answer to sub-question 5.

Chapter 6

Conclusion and discussion

6-1 Introduction

Having provided a detailed account of the implementation of affect in an active inference model. This research will be finalized with a discussion, conclusion and ideas that could be implemented as a follow-up to this research.

6-2 Discussion

The first and foremost point that should be highlighted, is that the final implementation was successful in:

1. Inference of uncertainty through precision optimization
2. Need prioritization, including:
   (A) The prioritization of needs in relation to other needs.
   (B) The prioritization of needs in relation to contextual opportunities and restrictions.

The modulation of precision here steals the show. Especially in the second task it lines up neatly with Solms' definition as it simply uses Equation 3-1c to estimate the exteroceptive noise and base its behaviour on this.

For the first task, the precisions of the priors were determined by hardcoding and not using the precision optimization equation. As mentioned in the Chapter 5 this could be explained in two ways. The first is that in order to be able to have an implementation that can use the equation for making the food precision rise, a more complex hierarchical structure needs to be devised. The challenge here that needs to be overcome is that due to the way this precision optimization equation is set up it computes precision as a result of incoming prediction error rather than it can enforce a desired precision. In other words, it could be that the function is more useful to estimate sensory noise, than it is to "enforce a desired precision". This is a suitable topic for future research

Another way of thinking about this is that the brain is able to set precisions in two different ways: Either by inferring uncertainty (as is done in the exteroceptive modules) or by just setting the precisions to a desired value, such that associated prediction errors will be subordinate to the precision rather than the other way around. From a biological standpoint this could make sense. The brain is after all able to modulate the influence of neurons by spreading neuromodulators through the forebrain as explained in sections 3-2 and 3-3. This neuromodulatory mechanism could be a biological argument for setting precision according to wishes instead of letting them be subordinate to incoming noise or prediction errors. [20] May provide some insight into this as it discusses the difference between "utility" and "uncertainty" with respect to decision-making. Here utility could relate to manually setting precision to wishes, whereas "uncertainty" could relate to deriving precision values using the formula. For the purpose of this research, it would suffice to conclude that by making use of either hard-coding precisions or using the estimation formula it is possible to make the model prioritize its needs. The way these precisions are in the end devised in a biologically plausible manner is a suitable topic for future research.

Another point for discussion is that this research has not looked at the optimization close enough to conclude that the minimal free energy has been reached every time. That is,

6-3 Final conclusion

Using the research sub-questions as a backbone, this research has now completed a storyline that can answer the main research question. For answering the main research question, it was first necessary to clarify two key characteristics of affect, each relating to their own research sub-question. The first is to provide a definition of affect based on Mark Solms' work. Therefore, the first sub-question was formulated as follows:

1. What is affect in both a neurological and computational sense?

There are two characteristics that together define the affective process. First of all, the definition as provided in Section 3-2 could be summarised concisely in the following way: affective feelings are constituted by the constant inference of uncertainty. Computationally this is done by precision optimization. This inference of precision serves as a monitoring mechanism on how well an organism is maintaining its homeostatic preferences. The process of precision optimization determines the registered valence (positive or negative) of affect by the organism. Here increasing prediction error indicates increasing uncertainty and is registered as negative affect. Decreasing prediction error is registered as positive affect.

Secondly, precision plays a crucial role in the prioritization of homeostatic needs. That is, as there are uncountable homeostatic states that all need to be given their due, organisms need a way to prioritize certain needs over others. This is the second crucial aspect of affect and is associated with the second sub-question:

2. What is the function of affect in both a neurological and computational sense?

The key concept here is that in active inference precisions are used as a gain modulator on prediction error. Errors that are afforded a high precision have a larger impact on estimates than precisions with a lower precision. Within the affective theory of this research, it means that deviations from homeostasis, or (changes in) the prediction error with the highest afforded precision come on top as affective awareness to the organism. In other words, through precision determines tho what extent an organism is aroused by a prediction error. The answer to the second sub-question is thus that the affective mechanism as defined by Solms gives the organism the ability to prioritize needs. This form of organisation adds adaptive value concerning the organism's survival.

Prioritization has to be mandated by something. This can be specified in two ways. What determines that one specific homeostatic need is prioritized over another? This is where the term "context" comes into play. Context as defined by Solms'can be described in two ways: 1. needs in relation to other needs and 2. needs in relation to opportunity. The finding s above lead to the required simulation results and thus answering the third sub-question:

3. What simulation results are needed to prove the model works according to Solms definition

Here it can concluded that these requirements are:

1. Inference of uncertainty through precision optimization
2. Need prioritization, including:
   (A) The prioritization of needs in relation to other needs.
   (B) The prioritization of needs in relation to contextual opportunities and restrictions.

With the 2 key aspects of affect in place and the first three research questions are answered. Chapter 4 dove more into a concrete implementation of the model. Doing so, this chapter provided an answer to the fourth sub-question:

4. What design specifications does the model need to demonstrate affect successfully

Here it became clear that the model needed to adhere to 4 specific design requirements in its structure:

1. Hydar needs **multiple** interoceptive needs that can compete.
2. Hydar needs an **Interoceptive** system that can measure and prioritize interoceptive prediction errors.
3. Hydar needs **Exteroception** to prioritize these needs according to context.
4. Hydar needs **action**

This provides a theoretical overview of a potential affective active inference model. Chapter 4 proceeds by providing a detailed, computational implementation, which answers the fourth research sub-question.

5. Can this be showcased in a simulation?

After completion of the model description, Chapter 5 addresses the final sub-question by showcasing an implementation of all described above. Here the model, which can be regarded as an "Affective active inference" model under the assumptions posed by sub-question 1, proved to be able to perform the two prioritization tasks as defined and thus adhere to the criteria posed by sub-questions 2 and 3. This finally provides a set-up for an answer to the final research question:

- Can Solms' theory on affect be implemented in a continuous active inference Model?

Concerning this research question, it should first of all be stressed that regardless of the results, it is might still be a long shot to say whether affect in its fullest form can be implemented in an active inference model. After all, affect in real-life organisms is unimaginably more complex than the simulation provided in this research. It would therefore be unreasonable to state that this model is really "Affective", in the sense that humans and other complex organisms are.

However, what can be concluded from this research is that using the most fundamental definitions of affect as provided by Solms, with a key aspect being the inference of uncertainty using the precision optimization equation as provided in Chapter 3, it can be shown that an active inference model can indeed prioritize needs in relation to one another as well as in relation to context. Hydar's ability to do so indicates that these mechanisms can indeed be used for an agent to steer its own incentives appropriately to what it internally wants to dictate (needs vs needs) as well as on what is on offer exteroceptively (needs vs opportunities). Thus, the results provide support to the feasibility of Mark Solms' hypothesis on affect. This could imply that it might prove fruitful to look further into these principles when it comes to developing more adaptive active inference models.

Taking it one step further, it could be argued that it might be worth it to look further into these principles to create an "Affective active inference" model that could simulate our own affective mechanisms in a more detailed way. In a broader perspective, as far-fetched as it may sound, this could then be seen as an effort to create Artificial Intelligence agents with conscious mechanisms that are similar to our own.

Appendix A

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.linalg import toeplitz, cholesky, sqrtm, inv
import math as m

# Define time settings
T = 1000
dt = 0.005
t = np.arange(0, T, dt)
N  = t.size

F = np.zeros(N)
x = np.zeros(N)
x_dot = np.zeros(N)
y_temp = np.zeros(N)
y_temp_dot = np.zeros(N)
y_food = np.zeros(N)
y_food_dot = np.zeros(N)
mu_temp= np.zeros(N)
mu_food = np.zeros(N)
phi_temp = np.zeros(N)
phi_food = np.zeros(N)
mu_xtemp = np.zeros(N)
mu_xtemp_dot = np.zeros(N)
mu_xfood = np.zeros(N)
mu_xfood_dot = np.zeros(N)
u = np.zeros(N)
mu_v = np.zeros(N)
mu_v_dot = np.zeros(N)
x_int = np.zeros(N)


# precisions
#sensor 1
omega_temp_z0 = np.ones(N)*5
omega_temp_z1 = np.ones(N)*5
```

```

#sensor 2
omega_food_z0 = np.ones(N)*5
omega_food_z1 = np.ones(N)*5

#for experiment 1
# omega = np.ones(N)*5
# omega_2 = np.ones(N)*5
# omega_3= np.ones(N)*3

# start = 0.25
# end = 0.75
# omega_4 =np.concatenate((np.ones(round(start*N))*0.3, np.linspace(0.3,30,round((1-start-(1-end))*N)), np.ones(round((1-end)*N))*3))
# PI_w = np.diag([10,1]) # experiment 1


# # experiment 2
omega = np.ones(N)*5
omega_2 = np.ones(N)*5
omega_3= np.ones(N)*10
omega_4 = np.ones(N)*10
PI_w = np.diag([1,0.1])


#learning rates
alpha = 1
alpha_u = 1



#initiazation
mu_temp[0] = 1
mu_xtemp[0] = 1
mu_food[0] =  1
mu_xfood[0] = 1
x[0]= 1

v_exa = np.zeros(N)
v_exb = np.zeros(N)
v_ey_temp = np.zeros(N)
v_ey_food = np.zeros(N)

v_eps_1 = np.zeros(N)
v_eps_2 = np.zeros(N)
v_eps_3 = np.zeros(N)
v_eps_4 = np.zeros(N)



for i in np.arange(0,N-1):
    #generative process of external position
    x_dot[i] = u[i]
```

```
    x[i+1] = x[i] + dt*x_dot[i]



    #generative process of internal states
    x_int = x[i+1]
    x_hung = x[i+1]


    #sensory states
    y_temp[i] = x[i]
    y_temp_dot[i] = x_dot[i]
    y_food[i] = x[i]
    y_food_dot[i] = x_dot[i]

    # if i>0.6*N:
    #     omega_4[i] = 5

    #COMMENT NOISE FOR EXP 1
    #UNCOMMENT NOISE FOR EXP 2
    if i >0.01*N:
        y_temp[i] = x[i]# + np.random.randn(1)*0.2#* np.sin(m.pi/N*i)
        y_food[i] = x[i]# + np.random.randn(1)*0.2


    if i > 0.2*N:
        y_temp[i] = x[i] + np.random.randn(1)*0.5
        y_food[i] = x[i] #+ np.random.randn(1)*0.5#* np.sin(m.pi/N*i)

    if i >0.5*N:
        y_temp[i] = x[i] #+ np.random.randn(1)*0.5
        y_food[i] = x[i] + np.random.randn(1)*0.5

    # set precisioon matrices for iteration
    PIa_z = np.diag([omega_temp_z0[i],0.1*omega_temp_z0[i]])
    PIb_z = np.diag([omega_food_z0[i],0.1*omega_food_z0[i]])
    # PI_wa= np.diag([1,1])
    # PI_wb= np.diag([1,1])



    #internal temperature estimation
    phi_temp[i] = x_int# + np.random.rand(1)*0.1
    eps_1 = phi_temp[i] - mu_temp[i]
    temp_prior = 0
    eps_3 = mu_temp[i] - temp_prior


    #hunger estimation
    phi_food[i] = x_hung
    eps_2  = phi_food[i] - mu_food[i]
    hunger_prior = 2
    eps_4 = mu_food[i] - hunger_prior
```

```



    #external prediction errors
    e_xtemp = np.array([[mu_xtemp_dot[i]], [0]]) - (np.array([[-mu_xtemp[i]],[-mu_xtemp_dot[i]]]) + np.array([[mu_temp[i]],[0]]))
    e_xfood = np.array([[mu_xfood_dot[i]], [0]]) - (np.array([[-mu_xfood[i]],[-mu_xfood_dot[i]]]) + np.array([[mu_food[i]],[0]]))
    e_ytemp = np.array([[y_temp[i]],[y_temp_dot[i]]]) - np.array([[mu_xtemp[i]],[mu_xtemp_dot[i]]])
    e_yfood = np.array([[y_food[i]],[y_food_dot[i]]]) - np.array([[mu_xfood[i]],[mu_xfood_dot[i]]])

    #make vectors with prediction error for controlling
    v_exa[i] = e_xtemp[0]
    v_exb[i] = e_xfood[0]
    v_ey_temp[i] = e_ytemp[0]
    v_ey_food[i] = e_yfood[0]

    v_eps_1[i] = eps_1
    v_eps_2[i] = eps_2
    v_eps_3[i] = eps_3
    v_eps_4[i] = eps_4

    #external depth estimation update
    dFdmu_xtemp = np.array([[1,1],[0,1]]).T.dot(np.dot(PI_w,e_xtemp)) - np.dot(np.identity(2),np.dot(PIa_z,e_ytemp))
    dmu_xtemp = np.array([[mu_xtemp_dot[i]], [0]]) - alpha*dFdmu_xtemp
    mu_xtemp[i+1] = mu_xtemp[i] + dt*dmu_xtemp[0]
    mu_xtemp_dot[i+1] = mu_xtemp_dot[i] +dt*dmu_xtemp[1]



    #external depth estimation update
    dFdmu_xfood = np.array([[1,1],[0,1]]).T.dot(np.dot(PI_w,e_xfood)) - np.dot(np.identity(2),np.dot(PIb_z,e_yfood))
    dmu_xfood = np.array([[mu_xfood_dot[i]], [0]]) - alpha*dFdmu_xfood
    mu_xfood[i+1] = mu_xfood[i] + dt*dmu_xfood[0]
    mu_xfood_dot[i+1] = mu_xfood_dot[i] +dt*dmu_xfood[1]


    #Free Energy MOET NOG WORDEN VERBETERD!!!!!!!!
    #F[i] =   0.5*(e_xfood.T.dot(PI_w.dot(e_xfood))) + 0.5*(e_xtemp.T.dot(PI_w.dot(e_xtemp)) +  e_yfood.T.dot(PIb_z).dot(e_yfood)  + e_ytemp.T.dot(PIa_z).dot(e_ytemp) + eps_1*omega[i]*eps_1 + eps_2*omega_2[i]*eps_2 + eps_3*omega_3[i]*eps_3 + eps_4*omega_4[i]*eps_4 - np.log(np.linalg.det(PI_w)) - np.log(np.linalg.det(PIa_z)) - np.log(np.linalg.det(PIb_z)))

    #causal temp state update
    dFdmu_temp= np.array([[-1,0],[0,0]]).T.dot(np.dot(PI_w,e_xtemp)) + np.array([[omega_3[i]*eps_3],[0]] ) - np.array([[omega[i]*eps_1],[0]] )
```

```
    dmu_temp= -alpha*dFdmu_temp[0]
    mu_temp[i+1] = mu_temp[i] + dt*dmu_temp[0]



    #causal hunger state update
    dFdmu_food = np.array([[-1,0],[0,0]]).T.dot(np.dot(PI_w,e_xfood)) + np.array([[omega_4[i]*eps_4],[0]]) - np.array([[omega_2[i]*eps_2],[0]] )
    dmu_food = -alpha*dFdmu_food[0]
    mu_food[i+1] = mu_food[i] + dt*dmu_food[0]



    #Action updating
    u_dot = -np.array([0,1]).dot(PIa_z).dot(e_ytemp) -np.array([0,1]).dot(PIb_z).dot(e_yfood)
    u[i+1] = u[i] + alpha_u*dt*u_dot


    alpha_1 = 1
    alpha_2 = 1
    alpha_3 = 1
    alpha_4 = 1
    alpha_za = 1.5
    alpha_zb = 1.5
    alpha_w = 1

    # precision updating
    # dPI_wa = alpha_w*0.5*(inv(PI_wa) - np.diag(e_xtemp.flatten())**2)
    # PI_wa = PI_wa + dt*dPI_wa
    # dPI_wb = alpha_w*0.5*(inv(PI_wb) - np.diag(e_xfood.flatten())**2)
    # PI_wb = PI_wb + dt*dPI_wb

    domega = alpha_1*0.5*(1/omega[i] - eps_1**2)
    omega[i+1] = omega[i] + dt*domega

    domega_2 = alpha_2*0.5*(1/omega_2[i]- eps_2**2)
    omega_2[i+1] = omega_2[i] + dt*domega_2


    #COMMENT OMEGA 3 AND 4 FOR EXPERIMENT 1
    domega_3 = alpha_3*0.5*(1/omega_3[i] - eps_3**2)
    omega_3[i+1] = omega_3[i] + dt*domega_3

    domega_4 = alpha_4*0.5*(1/omega_4[i]- eps_4**2)
    omega_4[i+1] = omega_4[i] + dt*domega_4

    domega_temp_z0 = alpha_za*0.5*(1/omega_temp_z0[i] - e_ytemp[0]**2)
    omega_temp_z0[i+1] =  omega_temp_z0[i] +dt*domega_temp_z0

    domega_food_z0 = alpha_zb*0.5*(1/omega_food_z0[i] - e_yfood[0]**2)
    omega_food_z0[i+1] =  omega_food_z0[i] +dt*domega_food_z0
```

```
    #just so the graph wont show zero at the end
    phi_temp[i+1] = x[i+1]
    phi_food[i+1] = x[i+1]

#water line drawing
water_space = np.linspace(0,50,N)
water_line = 0.045*np.sin(water_space)

bottom = np.ones(N)*2


#plot 3

plt.figure()
# plt.plot(t,ya, label = "measurement of outside temp ya")
# plt.plot(t,yb, label = " measurement of outside food yb")
plt.plot(t,mu_xtemp, label = "estimate of outside temp mu_x_temp", color = "limegreen")
plt.plot(t,mu_xfood, label = "estimate of outside food mu_x_food", color = "orange")
plt.plot(t, x, label = "real depth x (=real food/temp)", linewidth = "2", color = "r")
plt.plot(t,water_line)
plt.plot(t,bottom, color = "brown", linestyle = ':' )
plt.axvline(x = 0.2*T, color = 'b',linestyle = '--' )
plt.axvline(x = 0.5*T, color = 'b',linestyle = '--' )
plt.gca().invert_yaxis()
plt.xlabel("Time")
plt.ylabel("Hydar Depth (m)")
plt.title("External states")
#plt.legend(loc = 'lower right')
plt.legend(loc = 'center right', bbox_to_anchor = (0.65,0.8))


plt.figure()
plt.plot(t, mu_temp, label = "internal temp estimate mu_temp" , color = "limegreen")
plt.plot(t, phi_temp, label = "internal temp measurement phi_temp", color = 'b')
plt.plot(t, mu_food, label = "internal hunger estimate mu_food" , color = "orange" )
plt.plot(t, phi_food, label = "internal hunger measurement phi_food", color = "r",linestyle = '--' )
plt.axvline(x = 0.2*T, color = 'b',linestyle = '--' )
plt.axvline(x = 0.5*T, color = 'b',linestyle = '--' )
plt.gca().invert_yaxis()
plt.xlabel("Time")
plt.ylabel("Interoceptive state value")
plt.title("Interoceptive states")
plt.legend(loc = 'center right', bbox_to_anchor = (0.65,0.7))
# plt.plot(mu_dot)
```

```
plt.figure()
plt.plot(t,1/omega_temp_z0, label =  "external y_temp", color = "
    limegreen")
plt.plot(t,1/omega_food_z0, label = "external y_food" , color =  "orange
    ")
plt.axvline(x = 0.2*T, color = 'b',linestyle = '--' )
plt.axvline(x = 0.5*T, color = 'b',linestyle = '--' )
plt.legend()
plt.title("Exteroceptive variances (Attention)")
plt.xlabel("Time")
plt.ylabel("Estimated variance")


plt.figure()
plt.plot(t,omega_temp_z0, label =  "external y_temp", color = "limegreen
    ")
plt.plot(t,omega_food_z0, label = "external y_food" , color =  "orange")
plt.axvline(x = 0.2*T, color = 'b',linestyle = '--' )
plt.axvline(x = 0.5*T, color = 'b',linestyle = '--' )
plt.legend()
plt.title("Exteroceptive precisions (Attention)")
plt.xlabel("Time")
plt.ylabel("Estimated precision")


plt.figure()
plt.plot(t, 1/omega, label = "internal temp sensor ", color = "limegreen
    ")
plt.plot(t, 1/omega_3, label = "internal temp prior")
plt.plot(t, 1/omega_2, label = "hunger sensor", color = "orange")
plt.plot(t, 1/omega_4, label = "hunger prior", color = "r")
plt.axvline(x = 0.2*T, color = 'b',linestyle = '--' )
plt.axvline(x = 0.5*T, color = 'b',linestyle = '--' )
plt.xlabel("Time")
plt.ylabel("Estimated variance")
plt.title("Interoceptive variances (Affect)")
plt.legend()


plt.figure()
plt.plot(t, omega, label = "internal temp sensor", color = "limegreen")
plt.plot(t, omega_3, label = "internal temp prior")
plt.plot(t, omega_2, label = "hunger sensor", color = "orange")
plt.plot(t, omega_4, label = "hunger prior", color = "r")
plt.axvline(x = 0.2*T, color = 'b',linestyle = '--' )
plt.axvline(x = 0.5*T, color = 'b',linestyle = '--' )
plt.xlabel("Time")
plt.ylabel("Estimated precision")
plt.title("Interoceptive precisions (Affect)")
plt.legend()
```

Bibliography

[1] Karl J. Friston, Jean Daunizeau, James Kilner, and Stefan J. Kiebel. Action and behavior: a free-energy formulation. *Biological Cybernetics*, 102(3):227–260, March 2010.

[2] Karl Friston, Jérémie Mattout, and James Kilner. Action understanding and active inference. *Biological Cybernetics*, 104(1):137–160, February 2011.

[3] Pablo Lanillos, Cristian Meo, Corrado Pezzato, Ajith Anil Meera, Mohamed Baioumy, Wataru Ohata, Alexander Tschantz, Beren Millidge, Martijn Wisse, Christopher L. Buckley, and Jun Tani. Active Inference in Robotics and Artificial Agents: Survey and Challenges, December 2021. arXiv:2112.01871 [cs].

[4] Mark Solms. The Hard Problem of Consciousness and the Free Energy Principle. *Frontiers in Psychology*, 9, 2019.

[5] Mark Solms and Karl Friston. HOW AND WHY CONSCIOUSNESS ARISES: SOME CONSIDERATIONS FROM PHYSICS AND PHYSIOLOGY.

[6] Mark Solms. *The hidden spring: a journey to the source of consciousness.* Profile Books, London, 2021.

[7] Karl Friston, James Kilner, and Lee Harrison. A free energy principle for the brain. *Journal of Physiology-Paris*, 100(1-3):70–87, July 2006.

[8] Christopher L. Buckley, Chang Sub Kim, Simon McGregor, and Anil K. Seth. The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81:55–79, December 2017.

[9] Mateus Joffily and Giorgio Coricelli. Emotional Valence and the Free-Energy Principle. *PLOS Computational Biology*, 9(6):e1003094, June 2013. Publisher: Public Library of Science.

[10] Casper Hesp, Ryan Smith, Thomas Parr, Micah Allen, Karl J. Friston, and Maxwell J. D. Ramstead. Deeply Felt Affect: The Emergence of Valence in Deep Active Inference.

Neural Computation, 33(2):398–446, February 2021. Place: Cambridge Publisher: Mit Press WOS:000613887400004.

[11] Casper Hesp, Alexander Tschantz, Beren Millidge, Maxwell Ramstead, Karl Friston, and Ryan Smith. Sophisticated Affective Inference: Simulating Anticipatory Affective Dynamics of Imagining Future Events. In Tim Verbelen, Pablo Lanillos, Christopher L. Buckley, and Cedric De Boom, editors, *Active Inference*, Communications in Computer and Information Science, pages 179–186, Cham, 2020. Springer International Publishing.

[12] Donald Pfaff. *Brain arousal and information theory: Neural and genetic mechanisms*. Brain arousal and information theory: Neural and genetic mechanisms. Harvard University Press, Cambridge, MA, US, 2006. Pages: vi, 205.

[13] David J Chalmers. Facing Up to the Problem of Consciousness.

[14] Sigmund Freud. *Project for a Scientific Psychology (1950 [1895])*. Hogarth Press, 1966. Google-Books-ID: etztxQEACAAJ.

[15] Christophe Dupre and Rafael Yuste. Non-overlapping Neural Networks in *Hydra vulgaris*. *Current Biology*, 27(8):1085–1097, April 2017.

[16] Antonio Damasio and Gil B. Carvalho. The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews Neuroscience*, 14(2):143–152, February 2013. Number: 2 Publisher: Nature Publishing Group.

[17] Charel van hoof. Active inference code by example - 2.

[18] Karl Friston, Jérémie Mattout, and James Kilner. Action understanding and active inference. *Biological cybernetics*, 104(1-2):137–160, February 2011.

[19] Iris Hijne. Generalised Motions in Active Inference by finite differences: Active Inference in Robotics. 2020.

[20] Peter Dayan. Twenty-Five Lessons from Computational Neuromodulation. *Neuron*, 76(1):240–256, October 2012.