# Modeling Dyshomeostasis and Its Potential Causes by Active Inference

Raffael Theiler (10-928-893), Kaan Oktay (18-796-508), Oliver Kälin (19-945-336)

## I. Introduction

Mental disorders rank among the greatest burden of disease worldwide (Vigo et al., 2016). It has been estimated that in the European Union a third of the population suffers from mental disorders every year (Wittchen et al., 2011). However, treatment allocation is often based on trial and error because disease mechanisms still remain largely unknown. Identifying individual disease mechanisms has been introduced as a priority problem in psychiatry (Stephan et al., 2016a). This would allow for disease mechanism based differential diagnosis and pave the way for more effective treatment of mental disorders.

Petzschner et al. (2017) have proposed a framework for differential diagnosis in computational psychiatry (Figure 1). The framework builds on the idea that living beings need to keep bodily states e.g. physiological variables such as body temperature within a certain range to stay alive. Maintaining this homeostasis can be formulated as a feedback-control loop. In the framework, higher order hierarchies are added to this basic homeostatic reflex arc system. First of all, by adding a superior level which predicts future fluctuations the simple reactive system becomes pro-active. Secondly, a metacognitive level controlling the control capabilities of the system has been proposed. From this framework five potential causes of dyshomeostasis can be derived: 1) sensory input, 2) perception, 3) forecasting, 4) control action and 5) metacognition (Petzschner et al., 2017, Stephan et al., 2016b). The aim of this project was to simulate different causes of dyshomeostasis proposed in the framework of Petzschner et al. (2017).

Next, we introduce the concepts of homeostasis, dyshomeostasis and active inference. After that, we explain the research question in detail. In the remainder of the report we will then describe our simulations and results in detail.

### A. *Homeostatic Control as Bayesian Inference-Control Loop*

Homeostatic control is the process of keeping sensory signals of environmental states close to set-points (Stephan et al., 2016b). Stephan et al. (2016b) formulate homeostatic control from a Bayesian perspective. From this point of view homeostatic set-points represent expectations of prior beliefs about the state of hidden variables. The feedback arch of the control system is represented by Bayesian inference. Hidden states of the world (e.g. physiological variable x) have to be inferred from sensory input caused by the state of the world. In many cases Bayesian inference – updating beliefs about the hidden state of the world when new observations were made – have a general form. The belief updates are proportional to the prediction error which is the difference between the sensory input and the a priori expected sensory input, weighted by a ratio of prior and sensory input precision. Under this perspective the brain's objective can be described as finding the optimal model of the world in order to predict it optimally. In other words, brains try to be minimally surprised by their sensory inputs (Stephan et al., 2016b). Instead of updating its model of the world another option to achieve the same objective is by acting upon the environment which generates the sensory input. Taking action to reduce prediction errors in Bayesian inference is termed active inference (Stephan et al., 2016b). A prolonged difference between expected and actual sensory inputs yields to constant prediction errors and reflects a chronic state of surprise. Dyshomeostasis is therefore defined as chronic surprise (Stephan et al., 2016b).

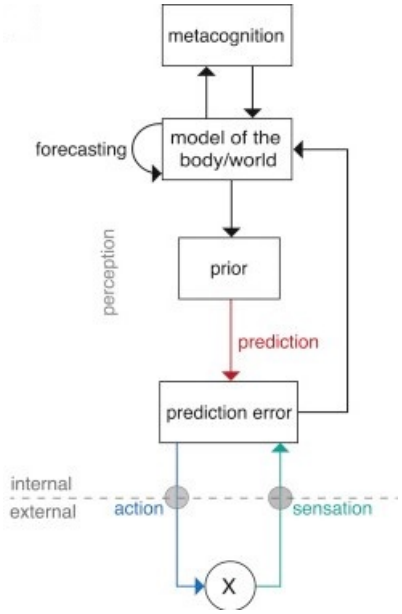


**Fig. 1:** Hiearachical Bayesian model of an inference-control loop proposed by Petzschner et al. (2017) (Figure is copied from Petzschner et al. (2017))

### B. *Inference-Control Loop as Hierarchical Bayesian Model and Causes of Dyshomeostasis*

The previously introduced inference-control loop can also be extended into a hierarchical Bayesian model (Petzschner et al., 2017). A first important extension is the introduction of allostatic control (Stephan et al., 2016b). It extends the simple reactive homeostatic control loop to become prospective. Allostatic control requires to acquire a model of the world that simulates the world state trajectory into the future ('forecasting') (Petzschner et al., 2017). This allows for changing prior

beliefs of states of the world in advance. Finally, at the top of the hierarchy a metacognitive level was proposed (Stephan et al., 2016b). It was suggested that the metacognitive level tracks the change of surprise of the lower levels and therefore is a representation of the belief about the control capabilities of the entire hierarchy (Stephan et al., 2016b).

The framework from Petzschner et al. (2017) (Figure 1) reveals distinct potential causes of the experience of dyshomeostasis (Petzschner et al., 2017, Stephan et al., 2016b). Dyshomeostasis can occur from lesions in sensory areas which is equivalent to erroneous feedback signals. Therefore, dyshomeostasis will be experienced irrespective of the correspondence of a homeostatic set-point and a true hidden (physiological) variable. Secondly, dyshomeostasis occurs when effector regions are lesioned. This results in the inability to take the necessary actions to reduce prediction errors. Clearly, dyshomeostasis can be caused by changes in the body itself which represents a disease that blocks homeostatic regulation (e.g. autoimmunological processes). Further, dyshomeostasis can originate from the level of allostatic control. Flawed forecasting of future bodily states results in maladaptive shifts of homeostatic priors. Lastly, it is possible that dyshomeostasis is caused by the brain itself. This can occur through flawed inference which corresponds to flawed perception, even though the sensory input is accurate. For example, overly rigid prior homeostatic beliefs about the states of the world result in not learning from observations and the occurrence of large prediction errors would be fought against with excessive actions. Another possibility of dyshomeostasis caused by the brain itself originates from metacognition. If the brain learns from past experiences that it cannot minimize surprise effectively with its actions or by learning to better predict sensory inputs, then the brain forms high-order beliefs about lack of control (Stephan et al., 2016b). These metacognitive beliefs about lack of control then lead to the subjective feeling of fatigue (Stephan et al., 2016b) and also down-regulate action taking as it does not improve the situation anyways.

### C. *Research Question*

A previous simulation of allostatic regulation of homeostatic control through active inference has been done in Stephan et al. (2016b). However, no explicit allostatic model has been specified in their simulation. Therefore, the simulation mainly incorporates a Bayesian homeostatic reflex arch, where the evolution of surprise and action are only dependent on the prediction error of the previous time point. However, for optimal action selection it is necessary to plan ahead. Planning allows to choose an optimal sequence of actions in order to reach one's goals by hypothetically running through all the possible consequences of each action. Roughly, this is what the active inference algorithm (Friston et al., 2009b, Sajid et al., 2019), which is based on the free energy principle (Friston, 2010), does. It finds optimal action sequences by minimizing future surprise over a time horizon. Here, optimal refers to a priori preferred states which correspond to the homeostatic set-points.

For this project we used this active inference learning algorithm to simulate homeostatic control similar to Stephan et al. (2016b). This brings the advantage that our simulation incorporates action planning over multiple time steps. Further, this allows us to incorporate allostatic control formally to the model. The aim of the project was to implement all levels of the hierarchical model proposed by Petzschner et al., 2017 and then perturb the system in order to simulate the different potential causes of dyshomeostasis previously described.

## II. Active Inference

### A. *Motivation*

As already indicated, the free energy principle states that all biological agents must minimize surprise about their sensory input (Friston, 2010, Stephan et al., 2016b). From the free energy principle active inference follows from the premise that action and perception aim at minimizing surprise (Friston, 2010). More formally, it is assumed that agents learn a generative model of how their sensory input has been generated. This is necessary because the true states of the world that generate observations (e.g. sensory input) are hidden to the agent. If we formulate the generative model as follows

$$P(\mathbf{o}, \mathbf{s}) = P(\mathbf{o}|\mathbf{s})P(\mathbf{s}) \tag{1}$$

then surprise can be formally defined as below:

$$S = -\log P(\mathbf{o}) \tag{2}$$

Intuitively, minimizing surprise corresponds to maximizing the evidence for the agent's generative model. Above, $P(\mathbf{s})$ is the agent's prior for the hidden states and $P(\mathbf{o}|\mathbf{s})$ is the likelihood of the outcome. We can obtain the probability of a given outcome $\mathbf{o}$ as follows:

$$P(\mathbf{o}) = \sum_{\mathbf{s}} P(\mathbf{o}|\mathbf{s})P(\mathbf{s}) \tag{3}$$

### B. *Variational Free Energy*

The model evidence $P(\mathbf{o})$ is often practically impossible to compute because the hidden state space can be extremely large. To solve this problem, we can utilize a tractable approximate distribution $Q(\mathbf{s})$ for the hidden states as below (Sajid, Ball, and Friston, 2019):

$$\begin{aligned} -\log P(\mathbf{o}) &= -\log \sum_{\mathbf{s}} P(\mathbf{o}, \mathbf{s}) \\ &= -\log \sum_{\mathbf{s}} Q(\mathbf{s}) \frac{P(\mathbf{o}, \mathbf{s})}{Q(\mathbf{s})} \\ &\leq -\sum_{\mathbf{s}} Q(\mathbf{s}) \log \frac{P(\mathbf{o}, \mathbf{s})}{Q(\mathbf{s})} \\ &= \sum_{\mathbf{s}} Q(\mathbf{s}) \log \frac{Q(\mathbf{s})}{P(\mathbf{o}, \mathbf{s})} \end{aligned} \tag{4}$$

where Jensen's inequality is used in the third step. As can be seen above, the last equation is an upper bound to the surprise. This upper bound is also called variational free energy ($F$) or (negative) evidence lower bound (ELBO). It can be rearranged as below:

$$\begin{aligned} F &= \sum_{\mathbf{s}} Q(\mathbf{s}) \log \frac{Q(\mathbf{s})}{P(\mathbf{o},\mathbf{s})} \\ &= \sum_{\mathbf{s}} Q(\mathbf{s}) \log \frac{Q(\mathbf{s})}{P(\mathbf{s}|\mathbf{o})P(\mathbf{o})} \\ &= \sum_{\mathbf{s}} Q(\mathbf{s}) \log \frac{Q(\mathbf{s})}{P(\mathbf{s}|\mathbf{o})} - \sum_{\mathbf{s}} Q(\mathbf{s}) \log P(\mathbf{o}) \\ &= \sum_{\mathbf{s}} Q(\mathbf{s}) \log \frac{Q(\mathbf{s})}{P(\mathbf{s}|\mathbf{o})} - \log P(\mathbf{o}) \\ &= D_{KL}\big[Q(\mathbf{s})||P(\mathbf{s}|\mathbf{o})\big] - \log P(\mathbf{o}) \end{aligned} \tag{5}$$

where $D_{KL}[\cdot]$ is the KL-divergence between two probability distributions. Therefore, surprise is minimized when the term $F$ is minimized because the KL-divergence cannot be negative. Further, the variational free energy is minimized when the similarity between the approximate variational distribution $Q(\mathbf{s})$ and the true posterior $P(\mathbf{s}|\mathbf{o})$ is maximized. These two separate objectives contribute to the optimization of variational free energy.

### C. *Expected Free Energy*

Free energy can be used to approximate a posterior about the hidden states while minimizing the surprise about the outcomes. However the formulation so far only allows us to do inference about the hidden states of the world but it does not give a tool for taking actions. To embed actions into free energy, we have to first introduce policies, which is a sequence of actions, as follows (Sajid et al., 2019):

$$u_t = \pi(t) \tag{6}$$

where $u_t$ is the action at time $t$ and $\pi$ is the policy over the time. We can reformulate the variational free energy to be dependent on time and policy as below:

$$\begin{aligned} F(t,\pi) =& \sum_{\mathbf{s}_t^\pi} Q(\mathbf{s}_t|\pi) \log \frac{Q(\mathbf{s}_t|\pi)}{P(\mathbf{o}_t,\mathbf{s}_t|\mathbf{s}_{t-1},\pi)} \\ =& \sum_{\mathbf{s}_t^\pi} Q(\mathbf{s}_t|\pi) \log \frac{Q(\mathbf{s}_t|\pi)}{P(\mathbf{o}_t|\mathbf{s}_t)P(\mathbf{s}_t|\mathbf{s}_{t-1},\pi)} \\ =& \mathbf{E}_Q\big[\log Q(\mathbf{s}_t|\pi) - \log P(\mathbf{o}_t|\mathbf{s}_t) \\ & - \log P(\mathbf{s}_t|\mathbf{s}_{t-1},\pi)\big] \end{aligned} \tag{7}$$

where $\mathbf{s}_t^\pi$ represents the states expected under the given policy $\pi$ and $Q$ represents the distribution $Q(\mathbf{s}_t|\pi)$. Using the time and policy dependent variational free energy definition above, the expected free energy is calculated by taking the expectation of $F(t,\pi)$ under the posterior predictive distribution $P(\mathbf{o}_t|s_t)$ and then summing over time as below:

$$G(t) = \sum_t G(t,\pi) \tag{8}$$

$$\begin{aligned} G(t,\pi) &= \sum_{\mathbf{s}_t^\pi,\mathbf{o}_t} P(\mathbf{o}_t|s_t)Q(\mathbf{s}_t|\pi) \log \frac{Q(\mathbf{s}_t|\pi)}{P(\mathbf{o}_t,\mathbf{s}_t|\mathbf{s}_{t-1},\pi)} \\ &= \sum_{\mathbf{s}_t^\pi,\mathbf{o}_t} P(\mathbf{o}_t|s_t)Q(\mathbf{s}_t|\pi) \log \frac{Q(\mathbf{s}_t|\pi)}{P(\mathbf{s}_t|\mathbf{o}_t,\mathbf{s}_{t-1},\pi)P(\mathbf{o}_t)} \\ &= \sum_{\mathbf{s}_t^\pi,\mathbf{o}_t} P(\mathbf{o}_t|s_t)Q(\mathbf{s}_t|\pi) \log \frac{Q(\mathbf{o}_t|\pi)}{P(\mathbf{o}_t|\mathbf{s}_t)P(\mathbf{o}_t)} \\ &= \sum_{\mathbf{s}_t^\pi,\mathbf{o}_t} P(\mathbf{o}_t|s_t)Q(\mathbf{s}_t|\pi) \log \frac{Q(\mathbf{o}_t|\pi)}{P(\mathbf{o}_t)} \\ &\quad - \sum_{\mathbf{s}_t^\pi,\mathbf{o}_t} P(\mathbf{o}_t|s_t)Q(\mathbf{s}_t|\pi) \log P(\mathbf{o}_t|\mathbf{s}_t) \\ &= \sum_{\mathbf{s}_t^\pi,\mathbf{o}_t} Q(\mathbf{o}_t,\mathbf{s}_t|\pi) \log \frac{Q(\mathbf{o}_t|\pi)}{P(\mathbf{o}_t)} \\ &\quad - \sum_{\mathbf{s}_t^\pi} Q(\mathbf{s}_t|\pi) \sum_{\mathbf{o}_t} P(\mathbf{o}_t|s_t) \log P(\mathbf{o}_t|\mathbf{s}_t) \\ &= \sum_{\mathbf{o}_t} Q(\mathbf{o}_t|\pi) \log \frac{Q(\mathbf{o}_t|\pi)}{P(\mathbf{o}_t)} \\ &\quad - \sum_{\mathbf{s}_t^\pi} Q(\mathbf{s}_t|\pi) H\big[P(\mathbf{o}_t|\mathbf{s}_t)\big] \\ &= \underbrace{D_{KL}\big[Q(\mathbf{o}_t|\pi)||P(\mathbf{o}_t)\big]}_{\text{expected cost}} + \underbrace{\mathbf{E}_Q\Big[H\big[P(\mathbf{o}_t|\mathbf{s}_t)\big]\Big]}_{\text{expected ambiguity}} \end{aligned} \tag{9}$$

where $P(\mathbf{o}_t|\mathbf{s}_t)Q(\mathbf{s}_t|\pi) = Q(\mathbf{o}_t,\mathbf{s}_t|\pi)$ and $Q = Q(\mathbf{s}_t|\pi)$. Prior preferences over the outcomes similar to the reward in reinforcement learning literature is defined as $P(\mathbf{o}_t) = P(\mathbf{o}_t|\pi)$. This result gives us an explanation what happens when we minimize the expected free energy. The first term, *expected cost*, minimizes the difference between expected outcomes under a policy and the prior outcome preferences. This can be interpreted as, the agent chooses policies that produce preferable outcomes. The second term, *expected ambiguity*, tries to minimize expected entropy of the generative model (likelihood of outcomes). This can be interpreted as the agent prefers a policy that increases the certainty and informativeness of its own outcome generation.

### D. *Overall Parametrization of Generative Scheme*

As mentioned, active inference concerns about two different generative schemes: The generative model of the agent and the generative process of the world. As in Sajid et al. (2019), we can parametrize the active inference problem using the tuple $(O,U,S,K,R,P,Q)$ where:

- $O$ is a finite set of outcomes.
- $U$ is a finite set of actions.
- $S$ is a finite set of hidden states.
- $K$ is a finite set of policies.

- $R(o, s, u)$ is a generative process of the world with outcomes $o \in O$ that are generated by hidden states $s \in S$ and actions $u \in U$.
- $P(o, s, \pi)$ is a generative model of the agent over outcomes, hidden states and policies $\pi \in K$ where $u_t = \pi(t)$.
- $Q(s, \pi) = Q(s_o|\pi) \dots Q(s_\tau|\pi)Q(\pi)$ is an approximate posterior over hidden states and policies.

Here, the generative process represents the transitions between hidden states in the world by actions and the generation of outcomes from these states which are observed by the agent. These transitions depend on the actions which are determined by the approximate posterior $Q$ about the hidden states. This approximate posterior is obtained by using the agent's own generative model which corresponds to the beliefs of the agent about the hidden states and policies. The overall model is depicted in Figure 2.

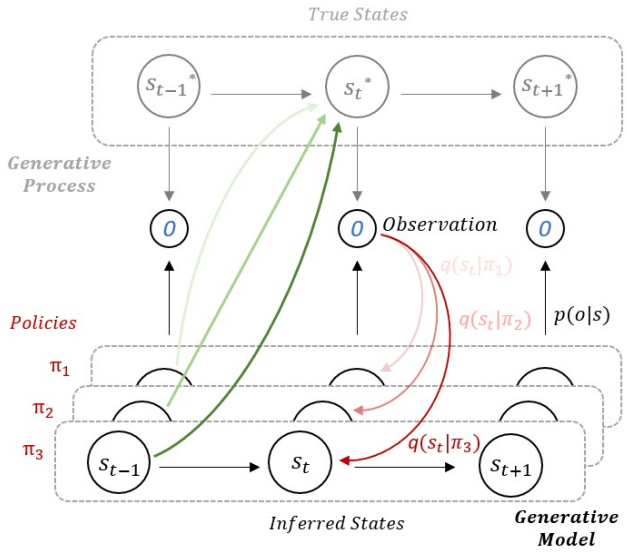


**Fig. 2:** Overall depiction of generative process and generative model used to infer the hidden states. The agent infers the true states of the generative process using the observations and its generative model (reproduced from: https://medium.com/@solopchuk/tutorial-on-active-inference-30edcf50f5dc, Solopchuk).

### E. *Optimization of Active Inference Objective*

Optimization in an active inference framework corresponds to updating the expected future hidden states under the policies, evaluating the expected free energy of each policy, taking Bayesian model averaging over the policies for the expected states and choosing the action which minimizes the expected free energy. Without going into the details of how optimization is performed computationally, we used a MATLAB library *SPM12* and specifically the file *spm_MDP_VB_X.m*, which uses variational message passing method for the optimization procedure.

### F. *Representation of Probability Distributions*

As previously mentioned, we used *SPM12* for the optimization procedure. This file requires some probability distribution tables to solve the active inference problem:

- **A**: A probability table for the generative model, $P(o_t|s_t)$. Each column in this table is a distribution over outcomes for a given hidden state.
- **B**: A stack of probability tables for state transitions, $P(s_t|s_{t-1}, \pi)$. For each possible action, there is a probability table where each column is a probability distribution over hidden states of the next step given the current hidden state.
- **C**: A vector of the agent's preferences (as log probabilities) over the outcomes, $P(o_t)$, similar to reward.
- **D**: A vector of the agent's prior probability distribution over the hidden states, $P(s_t)$, which is used in the beginning of the optimization procedure.

All these distributions are stationary. This means that they are the same for all time points $t$ and they do not change over a given policy. There is an optional input **s** to the optimization file. The variable **s** keeps the true hidden state ($s^*$). This is useful for numerous reasons that will be discussed in later parts. Lastly, there is another optional input $V$ to the file, which includes the policies the agent explores to find optimal actions and update its beliefs.

## III. Modeling of Homeostasis and Allostasis as Active Inference Problem

As mentioned before, active inference is a way of reducing the gap between an agent's prior expectations about observations and the actual observations at a given time point. Because the active inference framework utilizes both, changing perception and taking actions to achieve the objective of surprise minimization, it is a natural fit for modeling homeostatic and allostatic control. We devised two models with different perspectives on space-time. First, we introduce a context learning setting using an active inference agent that inspired our models. Then, we explain our models in detail.

### A. *Context Learning*

Contextual learning is best explained by an example. Consider a small, two dimensional, maze-like environment in which an agent must avoid holes and collect rewards, such as food. The agent is allowed to move up, down, left or right. This environment is generally known as frozen lake in the reinforcement learning literature Brockman et al. (2016). It's state space consists of all the fields of the maze with an assigned reward which represents a field with food, a hole or a neutral field. The frozen lake problem was solved using the active inference framework (AIF) in Sajid et al. (2019). The authors encode the maze's context, which is the position of food and holes, as a factor with its own hidden states in addition to the hidden states of the maze field. This corresponds to a combinatorial extension of the overall state space to include each possible maze configuration ahead of computation. Therefore the algorithm is fully aware of the entire environment before it finds an optimal strategy to navigate.

Given a relatively small world where a policy leads the agent from start to its terminal state, this is a feasible strategy. In each episode the entire game is played, therefore the agent learns a policy that covers the whole path from the start to an absorbing state in $T_p$ time steps. The agent learns by refining its posterior over states across many episodes during a trial. A posterior as we know it from the frozen lake example is optimal under active inference if the environment, i.e. the maze and the reward configuration, does not change and the agent remains inside of that known world. This strategy however reaches its computational limitation as soon as the state space grows to be large, either in cases where the agent tries to solve a continuous problem and/or when the world is too complex. It is also an infeasible strategy for partially known worlds where the agent discovers new states in different episodes by moving forward.

We ask ourselves if active inference can be used to navigate environments with either an infinite time horizon or a continuous state space or both, and to what degree. This is in particularly of our interest because many bodily signals, which can be mapped to the AIF's true state $s$, are continuous and without a time horizon. For that, we consider policies to be short term strategies in a (potentially indefinite) process.

This thought requires our experiments to consider two time frames. The first frame is assigned to the agent's episode and determines its short-term policy length $T_p$, similar to the frozen lake example. In our experiments, we usually assume $T_p \in [3, 10]$. It defines how far ahead of time the agent is able to predict future states $q(s|\pi)$ given a frozen episodal context. This can be seen as short-term allostatic control.

The second time frame is assigned to a trial, a set of consecutive episodes of total length $T_e$, unlike in the frozen lake experiment where $T_p = T_e$. We consider $T_e$ to be much larger than $T_p$. In that setting, the agent sees only a part of the whole trial in an active inference step. This introduces a different context, here called episodal context, depending on the progress $t_e$ in the trial $T_e$ that changes as soon as the agent moves forward. We propose two models that handle this episodal context differently. First, we perform contextual learning where a factor should mirror and understand the episodal context. In the second model, the agent has no direct representation of episodal context. It encodes episodal context directly by adjusting its control state to cope with the different influential forces given by context changes over a trial. Our goal is to observe how the agent chooses actions to maximize its utility during episodes that may randomly switch context to, for example prolonged (longer than one episode) episodes of stress (perturbations).

### *B. Model 1: Discrete Space-Time Modeling of Homeostasis and Allostasis by Active Inference*

Our first model interprets homeostatic and allostatic control as a context learning problem. We assume a scenario where the agent tries to keep homeostasis of a fictitious bodily state $x$ via its viscerosensory signals $y$. In addition, we provide an allostatic control mechanism using the context learning scheme. The agent's main aim is to keep the level of bodily state $x$ close to its homeostatic setpoint of the bodily state. To simulate it, we used two different factors that are hidden from the agent. This means that their have to be predicted by inverting the agent's generative model using the outcomes (viscerosensory signals) $y$.

*1) **Hidden factors**:* We call the hidden factors $f_1$ and $f_2$ and their states $s_{f_1}$ and $s_{f_2}$. The first factor represents the level of bodily state $x$ (e.g. physiological variable). It can take discretized values $s_{f_1} \in [-15, \dots, 15]$. In addition, the second factor defines the context of the world that the agent lives in and has the same state levels as the first one $s_{f_2} \in [-15, \dots, 15]$. These values from $-15$ to $15$ can be interpreted as relative values with respect to the normal level of a respective hidden state. For example, if we assume that the bodily state is body temperature and its normal is $99^o$F, then $s_{f_1} = 7$ means the body temperature is $106^o$F. For the context states, $s_{f_2}$ values can be interpreted as the most rewarding (preferable) value for the agent's bodily state $x$ e.g. a desired set-point for the body temperature. For example, while we are sleeping, the preferred body temperature decreases, then the set-point is adjusted accordingly and as a result, any external perturbations to the body temperature sensed via viscerosensory signals would result in a homeostatic control action.

*2) **Generative model for outcomes**:* Outcomes or observations, are the only quantities that the agent can directly observe. From outcomes the agent infers the hidden states using its own generative model. These outcomes are shared by both the generative process of the world and the generative model of the agent. For our model, we define $N = 31$ different outcomes which correspond to the total number of states for each hidden factor. Hidden states jointly map to the outcomes $o$ where $o \in \{1, \dots, 31\}$. When the true states of the hidden factors coincide, e.g. $s_{f_1} = s_{f_1} = 4$, the outcome is $o = 1$. When the true states differ from each other, the outcome is $o = |s_{f_1} - s_{f_2}| + 1$ e.g. $o = 7$ if $s_{f_1} = -1$ and $s_{f_2} = 5$. As mentioned before, the agent has prior preferences over outcomes which can be described as the agent's reward sensation when it observes a particular outcome.

*3) **Actions for state transitions**:* In the active inference framework, actions are vital to the agent to retain homeostasis. In this model, we use simple actions named *up*, *down* and *keep*. The other important point is that the agent's actions affect only the first hidden factor $f_1$. The first action *up* increases the level of the hidden factor by 1 e.g. from $s_{f_1} = 2$ to $s_{f_1} = 3$ while the action *down* decreases the level by 1 e.g. from $s_{f_1} = 2$ to $s_{f_1} = 1$. The last action "keep" does not change the level of the hidden factor and preserves its value.

*4) **Trials, episodes and policies**:* We can define a trial as consecutive episodes and a single episode as a time-frame in which different policies are evaluated by the agent. In this model, the agent can take three consecutive actions in each episode and we execute several episodes subsequently in a trial. When an episode ends, the posterior belief about the hidden states of $f_2$ and the true states of the hidden factors are

carried over to the next episode. This way, the agent chooses its optimal actions by considering the consequences of short-term policies (plans) and then retains its beliefs for the next episode. For each episode, we give the same policies to the agent with the following constraint: A policy either includes *up* or *down*, not both of them, and after a *keep* action, a different action is not allowed. By these rules, we assume that an agent prefers consistent actions. Therefore, the policies include all possible combinations of three actions without violating these constraints.

*5) **Allostatic control**:* In the active inference framework, the agent has an approximate posterior which represents its beliefs about the hidden states. In this model, we assume that an agent's belief about the hidden factor $f_2$ (context) is its set-point for the level of the bodily state $x$ represented by $f_1$. This set-point is then used for the homeostatic control of the bodily state $x$, or hidden state $f_1$. Overall, the allostatic control mechanism can be described as below:

- The true state of $f_2$ represents the context as defined previously. This true state cannot be seen directly by the agent but can be inferred via the observations.
- The beliefs of an agent about the state $s_{f_2}$ represents its prior about the bodily state $x$. When the context is changed externally, the agent updates its belief about the hidden state $s_{f_2}$ according to its reward sensation while executing its policies.

*6) **Homeostatic control**:* As mentioned, our posterior beliefs about $f_2$ is assumed to be the set-point for homeostatic control. Using this set-point, the agent takes an action on the bodily state $x$ which minimizes the expected free energy. This action results in a transition between hidden states of $f_1$ e.g. from $s_{f_1} = 2$ to $s_{f_1} = 3$ with action *up*.

### C. Model 2: Embedding Continuous Space-Time Dynamics into Active Inference

Model 2, even though implemented in the same framework as model 1, differs from it in three ways:

- A meta-cognitive factor is added.
- The agent's sensory inputs now originate from a side-car model in a continuous domain.
- The control problem is formulated to be undirected, which means the agent is not required to detect the need for up- or downregulation, solely the absolute distance between set-point $p$ and bodily state $v$, $|p-v|$ is relevant.

*1) **The undirected control problem (UCP)**:* The generative model $P$ of SPM12's AIF is limited to actions that are independent from each other. We are unable to specify $P(a_{f_i}|a_{f_k} = a_j)$. Therefore, any process that embodies different quantities $\mathbf{q} = (q_1, ..., q_n)^T$ influenced by the same action must be encoded in the same factor. This fact is important because our side-car model includes the previously introduced quantities $\mathbf{q} = (p, v)^T$.

E.g if we would model position $x$ and velocity $y$ of a vessel, then an increase of velocity $y$ would require a different set of actions for the position $x$ as this transition is directly affected by the vessel's current speed. Giving different sets of actions for position $x$ and velocity $y$ would allow the agent's generative process to sample actions that are not in line with the physical constraints of the vessel.

The exponential computational complexity of tabular active inference is a similar drawback compared to POMDPs. One strategy to overcome this issue is to use neural network approximators for densities Çatal et al. (2020). We fear that additional biases and effects introduced by neural networks may lead to wrong conclusions as they are only partially explainable. We choose to keep the tabular setting because it is sufficient to show our core ideas.

To implement the additional ideas for model 2 we had to simplify the directed control problem of model 1 considering these restrictions not only but also because of the hardware that was available for the experiments.

We define the undirected control problem here in the same setting as the directed one from model 1. The agent has to minimize the distance of the value $v$ (which is similar to the bodily state $x$ in model 1) with respect to the homeostatic set-point $p$. However the agent intrinsically knows the direction in which a correction is necessary, which is to move $v$ towards $p$. Without any additions, this would be a trivial problem to solve. The agent cannot deteriorate its position, no matter what action he takes. The key idea is that it is the agent's responsibility to decide whether a large but imprecise or a small but precise correction is necessary. This behavior is encoded in $B$ and depends on the magnitude $\kappa$. To simulate the action's precision, a noise term proportional to $\kappa$ is added.

*2) **UCP for continuous signals**:* Controlling bodily states is part of our research question. It was in our interest to control continuous space-time dynamics using the AIF because many bodily signals are understood as such. That way, we are able to compare the results to previous experiments in that domain. The AIF presented in SPM12 allows for discrete factors only. The discrete agent and the continuous part, which can be a model by itself, have to be strictly separated from each other to overcome this limitation. The SPM12 library contains exemplary material of classical control problems, such as mountain car (Friston et al., 2009a) or the video game DOOM (Cullen et al., 2018) that implement this separation in the AIF. In these examples, the true state $s \in S$ can be interpreted as true to the agent, $s$ does not necessarily represent the true state of the (continuous) world because it is an output of the agent. The authors remove this discrepancy with a mapping from continuous to discrete space and in reverse provide discrete actions that have an effect in continuous domain.

We use the same idea to map the bodily state to the discrete agent. We map the continuous signals $p \in \mathbb{R}$ and $v \in \mathbb{R}$ to 32 linearly distributed levels $\tilde{p} \in \tilde{P}$ and $\tilde{v} \in \tilde{V}$ each. These levels are encoded in a single factor $f_1$. Its hidden state is organized as a vector of length 1024 such that:

$$S_{f_1} = \left[ s^{\tilde{p}_1}_{\tilde{v}_1} \dots s^{\tilde{p}_1}_{\tilde{v}_{32}} \dots s^{\tilde{p}_2}_{\tilde{v}_1} \dots s^{\tilde{p}_2}_{\tilde{v}_{32}} \dots s^{\tilde{p}_{32}}_{\tilde{v}_{32}} \right]^T \tag{10}$$

In reverse, the 5 discrete actions $\tilde{a} \in \tilde{A}$ map to 5 different slopes $\kappa \in \mathbb{R}$, $\tilde{a}_i \mapsto \kappa_i$. $\kappa_i$ is then applied in continuous domain for a fixed amount of time. More formally, the outcome of an action $P(s_{t+1}|s_t\pi, a_t)$ is determined by a stochastic first order ODE:

$$\frac{df(p,v)}{dt} = \kappa(p-v) + \epsilon \tag{11}$$

where $p$ is a reachable set-point and $v$ is the current value of the bodily state. The equality includes a noise term $\epsilon$ with variance proportional to the slope $\kappa$:

$$\epsilon \sim N\left(0, \frac{\kappa}{2}\right) \tag{12}$$

During an episode, after each active inference step, we use a linearized implementation of the ODE to evaluate the outcome of $\kappa_i$ of that step's action $\tilde{a}_i$ in continuous domain for a fixed amount of time $\Delta t$ and then apply a mapping $d : [p, v] \mapsto [\tilde{p}, \tilde{v}]$ back to discrete space. This outcome is then used in the next active inference step.

For simplicity, we kept the continuous side-car model in its most basic form. We observed that in the SPM12's AIF the outcome generating process may also be directly embedded in $R$, which is inside the frameworks main function. There are several extensions for other projects, however such a change seemed out of scope for this project.

*3) **Parametrization of the active inference agent**:* In the previous section, we introduce a stochastic ODE as an extension to model 1. In model 2 however, we also adjust the probability tables $A, B, C, D$. Initially, we give uniform priors for $D$.

Using a linearized implementation of the ODE for a fixed time $\Delta t$, we are able to simply sample probabilistic transitions in between states for all the actions $\tilde{A}$ (using the respective $\kappa$). The linearized ODE can be seen as a function $b$:

$$b : [p, v]^{f_1}_t \mapsto [p, v]^{f_1}_{t+\Delta t} \tag{13}$$

Using $b$, we define the transition probabilities $P(s_t|s_{t-1})$, required for $B$, for each action $\tilde{a} \in \tilde{A}$, given an inverse mapping of $d$, $c : [\tilde{p}, \tilde{v}] \mapsto [p, v]$ from discrete to continuous space:

$$p^{kl}_{ij} = P\left(s^{\tilde{p}_i}_{\tilde{v}_j} | b, c\left([\tilde{v_k}, \tilde{p_l}]^{f_1}_{t-\Delta t}\right)\right) \tag{14}$$

where $p^{kl}_{ij}$ contains probability mass proportional to the distance:

$$|c(\tilde{p_i}, \tilde{v_j}) - b(c(\tilde{p_l}, \tilde{v_k}))| \tag{15}$$

and is zero, if $\tilde{v_j}$ and $\tilde{p_i}$ are not in the direct neighborhood of $v_{t+\Delta t}$ and $p_{t+\Delta t}$. i.e. if $b$ does not move us from position $kl$ close to $ij$ in the next time step $t + \Delta t$, then we do not give any probability mass for $I = L = |\tilde{P}| = 32$ and $J = K = |\tilde{V}| = 32$.

We define the stack of probability tables for state transitions $B$ as follows:

$$B_{\tilde{a}_k} = \begin{bmatrix} b^{11}_{11} & b^{12}_{12} & \dots \\ \vdots & \ddots & \\ b^{11}_{JK} & & b^{KL}_{IJ} \end{bmatrix} \tag{16}$$

We define the outcome $O_1$ in the algorithms probability table $C$ as a distance from zero to $max(\{\forall i, j : d_{ij} = |p_i - v_j|\})$. The smaller that distance, the better the outcome. As for every other quantity of the AIF, $O_1$ is also tabular and set to have five different levels.

The likelihood in $A$ for an outcome $o^1_i \in O_1$ for a state $s^{\tilde{p}_i}_{\tilde{v}_j}$ depends on the actual distance $\delta v = |p - v|$. We use a normal distribution $N(\delta v, o^1_i, \sigma^2)$ with mean $o^1_i$ and variance $\sigma^2$ to determine the probability of outcome $o^1_i$ given the current distance.

*4) **A metacognitive factor**:* We propose a simple extension to arrive at an agent with meta-cognitive capabilities in Figure 3. A second factor $f_2$ encodes the agent's belief about its self-efficacy. It is added to the previously introduced undirected control problem. Its binary states $S_{f_2} = \{s_c, s_h\}$ are interpreted as the agent either being self-confident $s_c$ or helpless $s_h$. A new outcome modality $O_2 = \{o_-, o_+\}$ is tightly coupled to this factor. It models a large reward that can easily be consumed by the agent $o_+$, independent of the actual control state, we call it the compensatory satisfaction. The other state $o_-$ yields a low amount of reward.

The AIF allows us to recognize very suitable analogies in between parts of the model and psychological circumstances. In the scope of this project, the outcome modality $O_2$ may be interpreted in a general sense as a quantity of satisfaction gained from one of many compensatory mechanisms that humans exhibit when under-performing. During one trial, the agent is exposed to different phases of stress, i.e. strong external perturbation of the regulated quantity. Given two sources of reward, $O_1$ and $O_2$, the agent may balance its income optimally under AIF optimization, which may entail neglecting the primary task in favour of the compensatory reward. On the other hand, transitions from and to the helpless state are modeled. This allows for a configurable short term expected tolerance to stress as well as an adjustable expected recovery rate.

*5) **Parametrization of the metacognitive factor**:* As with $f_1$, $f_2$ requires a formal introduction of the tables $A$ and $B$ and its actions. We interpret $B$ as the agents tendency to go into helpless state, and its capability to recover from this condition. With this assumptions $f_2$ should know three different actions: Become helpless $a_{a \to h}$, become active $a_{h \to a}$ and stay in current state $a_s$.

Using probabilities for "becoming helpless" $p_h$ and "being able to recover $p_r$ we define the following set of $B$:

$$B_{a_{a\to h}} = \begin{bmatrix} 1-p_h & 0 \\ p_h & 1 \end{bmatrix} \tag{17}$$

$$B_{a_{h\to a}} = \begin{bmatrix} 1 & p_r \\ 0 & 1-p_r \end{bmatrix} \tag{18}$$

and $B_{a_s} = I_2$ as the identity matrix.

The outcome likelihood in $A^{O_k}$ for any outcome $O_k$ is defined as $P(o_k|f_1, f_2)$, thus adding a second factor requires us to modify $A^{O_1}$ as well. To incorporate helplessness in $f_1$ we add a pseudo state to $O_1$ which has a very low preference (reward). As soon as $f_2$ is in helpless state, it forces the outcome likelihood of $A^{O_1}$ to concentrate on that pseudo state, effectively ignoring any reward that is bound to the control task.

For the second outcome $O_2$ we define the corresponding $A^{O_2}$ such that the agent decides to not use the compensatory reward in active (confident) state ($s^{f_2} = s_c$) meaning that we give a low likelihood for $o_+$ uniformly for any state of $f_1$ and high likelihood for $o_-$. We reverse this configuration as soon as the agent decides to be helpless ($s^{f_2} = s_h$)

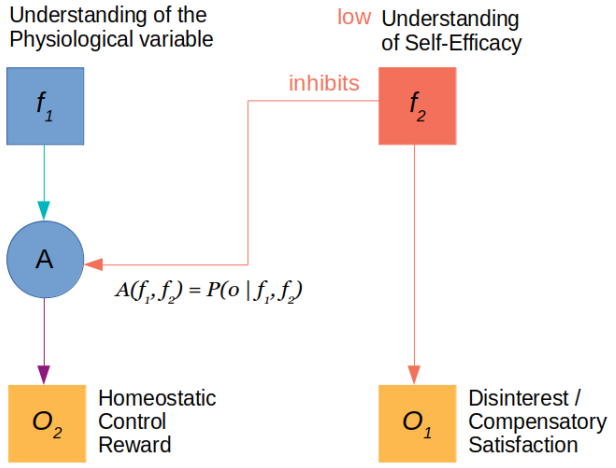


**Fig. 3:** A schematic of the causal relations for factors and outcomes of model 2. $f_1$ infers on the current state of the undirected control problem. $f_2$ is a meta-cognitive factor that monitors the agent's capability of adapting to the current level of stress (context). The agent incorporates two outcome modalities, one which favours the optimal solutions of the undirected control. The other modality has no specific preference, i.e is uniformly distributed over outcomes. In that setting, we observe the agent to balance its preference for outcomes, seen as rewards, based on the difficulty of the current context. In helpless situations, $O_2$ is priorized over $O_1$.

## IV. Experiments and Results

### A. *Modeling Potential Causes of Dyshomeostasis*

In this part, we present the results from the simulation of an active inference agent which can perform homeostatic and allostatic control. To illustrate this process, we should observe how the active inference agent and its hidden states react to given perturbations. All the experiments in this part are performed by model 1 as follows. We change the true state ($s_{f_2}$) of the hidden factor $f_2$ (context state) by $+15$ or $-15$ at certain time-points ($t = 61, 181, 301, \ldots, 541$) by hand to observe allostatic control (and homeostatic control right after that). In addition, we perturb the true state ($s_{f_1}$) of the hidden factor $f_1$, also named as bodily states $x$, by random values from $-13, -12, -11, -10, 10, 11, 12, 13$ at certain time-points ($t = 16, 136, 256, \ldots, 496$) to observe the homeostatic control mechanism of the agent. In order to describe the potential causes of dyshomeostasis in the context of the active inference agent, we should first define an agent with ideal properties as follows:

- The actions of the agent are deterministic which means each action does its intended transition with probability 1 e.g. when $s_{f_1} = 4$ is given and the agent takes action *up*, then it always ends up $s_{f_1} = 5$.
- The generative model of the agent is also deterministic which means that the probability distribution over the outcomes for given hidden states $s_{f_1}$ and $s_{f_2}$ is concentrated on a single outcome with probability 1 e.g. when $s_{f_1} = 2$ and $s_{f_2} = 5$ are given, then outcome is always $o = 4$.
- The agent has exponentially higher preference (high reward) as the level of the outcome $o$ decreases e.g. when the outcome goes from $o = 3$ to $o = 2$, the log-magnitude of preference increases by 1.

Using these ideal properties, we simulated an agent. The result can be seen in Figure 4. As can be seen, the agent can almost perfectly perform its homeostasis and allostasis regulation. Whenever the level of context changes, it adjusts its homeostatic set-point according to the preferences updated by the new context. Furthermore, when we perturb the bodily state externally, our agent can achieve the homeostatic control successfully.

*1) **Faulty sensory information**:* In this part, we simulate an active inference agent with faulty sensory information which means that the mapping from hidden states to the outcome is not perfect. As opposed to the agent with ideal properties, the likelihood over outcome levels is a probability distribution given hidden states and it decreases exponentially as the outcome level gets further away from the ideal case, similar to a Gaussian distribution with a mean at the ideal outcome state. This may cause an illusion of dyshomeostasis and homeostasis for the agent. By changing the variance of the Gaussian, we can model an agent with different levels of faulty sensors. The effect of faulty sensors on the homeostatic and allostatic control is minimal for low variance, as can be seen in Figure 5. However with higher variance in Figure 6, the allostatic and homeostatic regulation of the agent exhibit minor fluctuations around their ideal levels.

*2) **Abnormal motor functions**:* In this part, we simulate an active inference agent with abnormal motor functions which means that the actions of the agent are not perfect and may result in unintended transitions. In contrast to the agent with

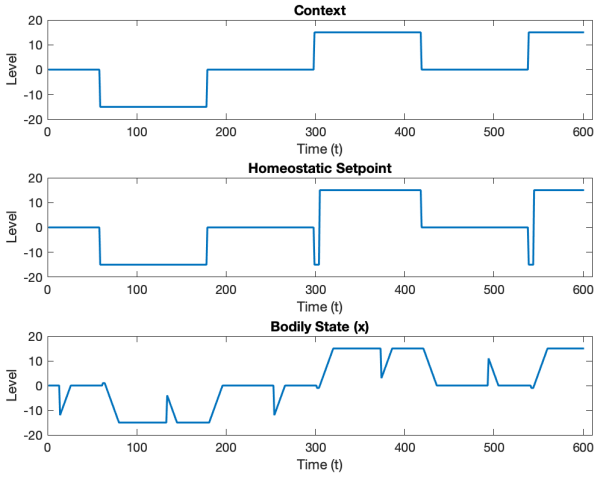


**Fig. 4:** Simulation of an agent with ideal properties. In the **top** plot, the change in context at the given time points can be seen. After the context change, the homeostatic set-point follows the context quickly as depicted in the **middle** plot. In the **bottom** plot, perturbations and the homeostatic regulation on bodily states $x$ can be seen. After each perturbation, the bodily state is brought back to the homeostatic set-point by the actions of the agent.

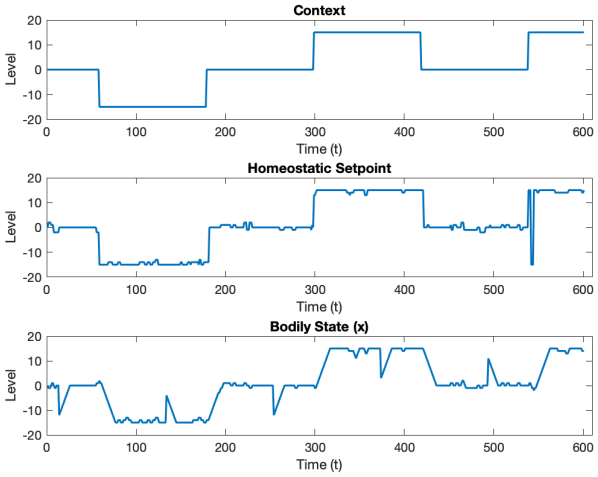


**Fig. 6:** Simulation of an agent with high-variance ($\sigma^2 = 1$) faulty sensory information. In the **top** plot, the context changes at certain points in time. After the context change, the homeostatic set-point follows the context with small fluctuations in the **middle** plot. In the **bottom** plot, after the perturbations on bodily states $x$ occurred or the set-point changed, the homeostatic regulation brings the bodily state back to the set-point level with small deviations around the ideal level.

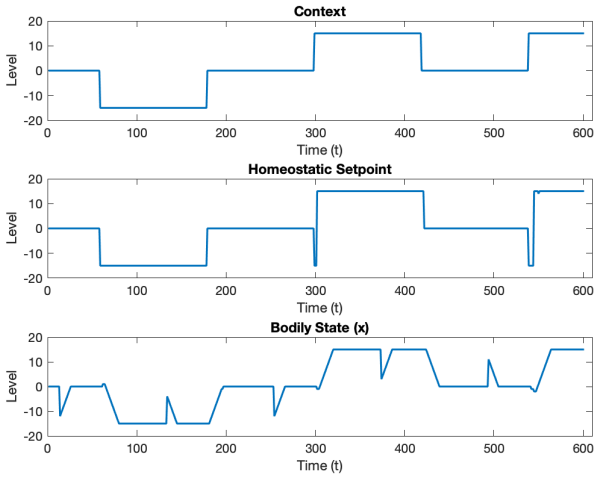


**Fig. 5:** Simulation of an agent with low-variance ($\sigma^2 = 0.5$) faulty sensory information. In the **top** plot, the context changes at certain points in time indicated by a step. After the context change, the homeostatic set-point follows the context successfully in the **middle** plot. In the **bottom** plot, after the perturbations on bodily states $x$ occurred or the set-point changed, the homeostatic regulation brings the bodily state back to the set-point level.

ideal properties, this agent has a probability distribution over the next state of hidden factor $f_1$ for a given current state and an action. This likelihood decreases exponentially as the hidden state level gets further away from the ideal case, similar to a Gaussian distribution with a mean at the ideal next state. As mentioned, this abnormality may result in unintended consequences of actions which can be catastrophic for the agent in terms of the homeostatic regulation. By changing the variance of the Gaussian, we can model an agent with different levels of motor function abnormality. As can be seen in Figures 7 and 8, the motor function abnormality can severely affect the homeostatic control ability of the agent when the variance is high while a low variance of abnormality only results in small perturbations on the homeostatic control of the bodily state. In contrast, the allostatic control is not affected by the motor function abnormalities.

*3) **Decreased reward sensitivity**:* In this part, we simulate an active inference agent with various levels of decreased reward sensitivity. In the ideal case, the preference for an outcome level exponentially increases as the outcome level gets closer to the most desired point determined by the context. In contrast to the ideal case, this agent cannot get the reward sense when it is too far away from its ideal level which means its preferences are the same after a certain distance to the most desired point. An agent with this property can be indifferent to any action in a region where all the states have the same preference and finding a successful policy to escape this region can be hard.

To simulate the effect of the indifference region (the region with indifferent states) size, we simulated two agents with bigger and smaller indifference regions. The agent with smaller indifference region gets the same small reward when it is further away from the ideal by more than 9 levels and the number of indifferent states is $n_d = 22$. For the other agent with bigger indifference region, this number is $n_d = 27$. As can be seen in Figures 10 and 9, the decreased

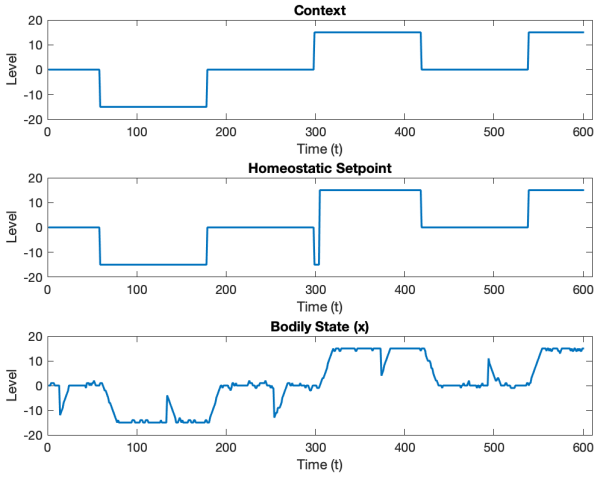


**Fig. 7:** Simulation of an agent with low-variance ($\sigma^2 = 0.5$) abnormal motor functions. In the **top** plot, the context changes at certain points in time. After the context change, the homeostatic set-point follows the context successfully in the **middle** plot. In the **bottom** plot, after the perturbations on bodily states $x$ occurred or the set-point changed, the homeostatic regulation brings the bodily state back to the set-point level and keep it there with insignificant fluctuations.

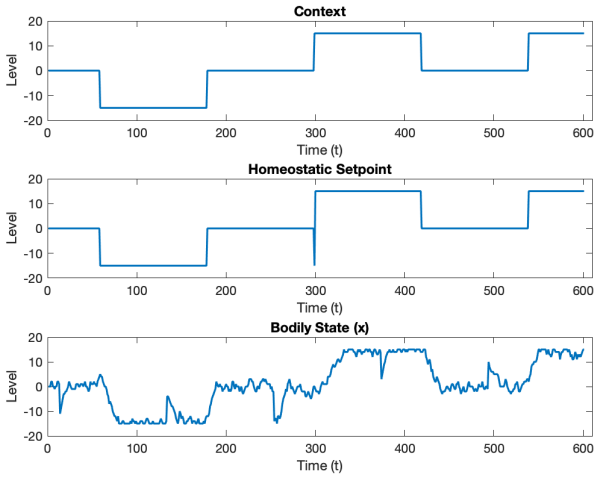


**Fig. 8:** Simulation of an agent with high-variance ($\sigma^2 = 1$) abnormal motor functions. In the **top** plot, the context changes at certain points in time. After the context change, the homeostatic set-point follows the context successfully in the **middle** plot. In the **bottom** plot, after the perturbations on bodily states $x$ occurred or the set-point changed, the homeostatic regulation brings the bodily state back to the set-point with long periods of significant deviations from the exact set-point.

reward sensitivity significantly harms the homeostatic control while the allostatic control works as in the ideal case. The agent cannot perform homeostatic regulation of the bodily state at the later time points, especially worse for the bigger indifference region case Figure 10. The agent sometimes cannot escape the indifference regions with the lack of proper reward signal.

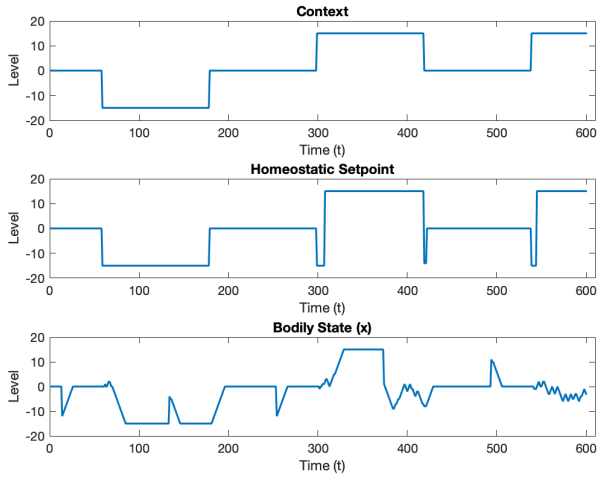


**Fig. 9:** Simulation of an agent with number of indifferent states $n_d = 22$. In the **top** plot, the context changes at certain points in time. After the context change, the homeostatic set-point follows the context successfully, sometimes with a small delay, in the **middle** plot. In the **bottom** plot, after the perturbations on bodily states $x$ occurred or the set-point changed, the agent experiences long periods of dyshomeostasis when it cannot escape its indifference region.

### *B. Simulation of Metacognitive Problems*

We simulated meta-cognitive self-efficacy with model 2. The agent's shortcomings should manifest themselves without any interference except stress, corresponding to to strong external perturbation, applied to the bodily state $v$. Opposed to the previous experiments, the agent shows an intrinsic weakness to continual stress given by $p_h$.

The experiment is initialized such that $\tilde{v} = 2$ and $\tilde{p} = 2$. The total duration of this experiment $T_e$ is 30 episodes. The policy length $T_p$ is 5 for a total of 150 time steps. One trial of the experiment includes five different episodic contexts:

1) 1-5: No perturbation, the signal is only influenced by the noise $\epsilon$.
2) 10-40: In this period, $v$ is perturbed strongly three times by immediately setting it to $\tilde{v} = 28$.
3) 30: A moderate perturbation of $v$ is performed ($\tilde{v} = 5$).
4) 60: A strong perturbation of $v$ is performed ($\tilde{v} = 28$).
5) 70-90: In this period, $v$ is perturbed strongly five times by immediately setting it to $\tilde{v} = 28$.
6) 90-150: A long period without perturbation.

Figure 11 shows the trajectories of the underlying continuous dynamical system $(p, v)_t$ and the perceived trajectory $(\tilde{p}, \tilde{v})_t$ of this experiment and summarizes the posterior over actions of the generative model. We observe several noticeable properties in the posterior distributions:

Initially ($t = [1, 5]$), the agent reacts to strong perturbations with a high probability of emitting action $\kappa_5$, which is the

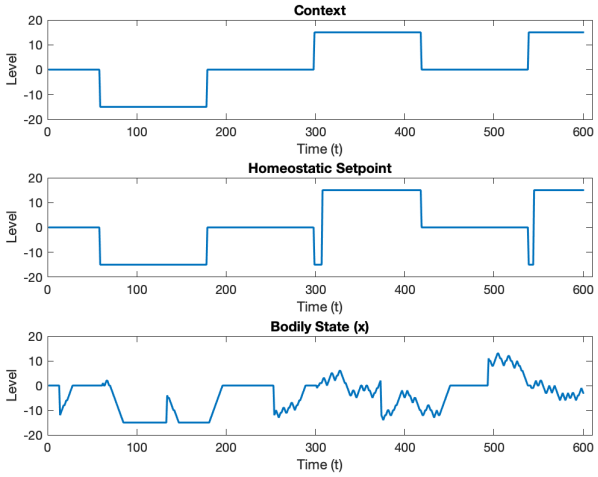


**Fig. 10:** Simulation of an agent with number of indifferent states $n_d = 27$. In the **top** plot, the context changes at certain points in time. After the context change, the homeostatic set-point follows the context successfully, sometimes with a small delay, in the **middle** plot. In the **bottom** plot, after the perturbations on bodily states $x$ occurred or the set-point changed, the agent experiences longer periods of dyshomeostasis (compared to the smaller indifference region case) because exponentially increasing reward signal (prior preferences) is restricted to a very small number of states.

expected behavior. A strong perturbation should be balanced out immediately. The included noise $\epsilon_{\kappa_5}$ that is proportional to the slope of action $\kappa_5$, hereby large, is not an issue because $v$ is far away from its set-point $p$. Therefore, the problem of overshooting the target $p$ because of $\kappa$ imprecision is almost 0.

If the agent endures prolonged stress (for $t$ in range $[10, 40]$ and $[70, 90]$) that pushes $v$ too far away from $p$ it loses its self-confidence about its ability of control. The agent signals to transition to a helpless state. This can be seen in the bottom plot of Figure 11, where action 1 corresponds to $a_{a \to h}$. Immediately after, the agent tries to recover. We can see this from the posterior belief over actions of $f_2$ which accummulates most of the probability mass on $a_{h \to a}$. It takes a moment to fully recovery due to the low recovery probability $p_r$.

If the agent transitions into the helpless state ($s^{f_2} = s_h$, for $t$ in range $[10, 40]$ and $[70, 90]$) we observe a causual relation in between two disabilities:

First, the agent loses its ability to infer on $f_1$'s hidden state $q(s^{f_1})$. This is a similar symptom as observed from the decreased reward sensitivity disease of model 1. The agent looses it's certaincy about the hidden state. However, in model 2 the agent transitions from and to this state based on the episodal context. Figure 12 shows that this is not an instantaneous change. The posterior slowly degenerates to a very flat distribution.

This is the cause for the second disability. With an uninformative $q(s^{f_1})$ the agent is also unable to show preference towards an optimal action $\tilde{A}^{f_1}$ which becomes apparent in Figure 11's posterior over actions of $f_1$ during time intervalls $[20, 40]$ and $[70, 90]$. In other words, the agent becomes indifferent about which actions to take from $\tilde{A}^{f_1}$.

As soon as the agent moves from stressful episodic contexts (1-5) towards the final (stress-free) context (6, $t = [100, 150]$) it adapts to interpret the $v$ signal differently. By looking at the agents perceived trajectory, in comparison to the posterior distribution over actions, we observe that the agent learns to choose the smallest feasible action $\kappa$ to offset incoming noise. The noise is now the relevant force moving $v$ away from $p$, that noise however is magnitudes smaller than the external perturbations during stressful episodes.

The purpose of the experiment presented in figure 13 is to have a baseline on how the agent behaves during a trial of complete helplessness. By changing the reward structure of $O_1$ we made it infeasible for the agent to choose any other state than ($s^{f_2} = s_h$) for $f_2$.

## V. Discussion

In perspective of the proposed framework for differential diagnosis in computational psychiatry by Petzschner et al. (2017), we summarize the key results from our models here. Our models showed that the AIF can be used to model homeostatic and allostatic regulation of an agent (i.e. a living organism). This framework offers various capabilities to model the agent's interaction with its environment.

In model 1 we simulate allostatic control by implementing the inference-control loop from Petzschner et al. (2017) in the AIF. Model 1 allowed us to simulate 2 distinct causes of dyshomeostasis proposed in (Petzschner et al., 2017). We simulated faulty sensory input and erroneous action taking. In addition, we simulated the case of decreased reward sensitivity. As seen in results, faulty sensory information with high variance may cause illusion of dyshomeostasis and result in the fluctuations in the allostatic control process while abnormal motor functions does not affect the allostasis. In contrast, abnormal motor functions with high variance of transition likelihood severely harm the homeostatic regulation process and cause significant perturbations around the exact set-point. By going backwards from these effects to the abnormalities, we can diagnose the cause of the failures in the agent's homeostatic and allostatic regulation process, which can be considered as *differential diagnosis*. Having this information, we can infer about the cause of perturbations if we know which process (allostasis or homeostasis) is disrupted. In addition, with the same variance, abnormal motor functions cause more severe disruptions to the homeostatic control than faulty sensors. We can assume that the long periods of sustained perturbations to the regulation of bodily states are more likely to be caused by the errors in motor functions.

Another capability which is very important for the agent's homeostatic and allostatic regulation is its preferences. This

corresponds to its reward mechanism. As shown with the experiments, the agent's preferences play an important role in its regulation process and the abnormalities can cause much more significant disruptions than the other abnormalities in the motor and sensory functions. This can be interpreted as the lack of proper reward signal may result in periods of sustained dyshomeostasis. While other abnormalities result in the perturbations around the homeostatic set-point for the bodily state, the agent experiences sustained dyshomeostasis in this case.

Stephan et al. (2016b) argue that if the agent cannot reduce dyshomeostasis by whatever the agent does, then the belief about lack of control will materialize. This belief about the level of mastery is represented by the meta-cognitive level. Model 2 aimed at modeling the meta-cognitive level. Stephan et al. (2016b) hypothesized that fatigue and potentially also symptoms of depression originate from this meta-cognitive level. The feeling of fatigue and helplessness arise, if any action one takes, does not help to restore homeostasis. Our model 2's second meta-cognitive factor learns exactly this relationship. However, it goes a step further, because the meta-cognitive factor effects action taking in other cognitive areas. Model 2 shows the characteristic that when the agent learns that actions can not restore homeostasis, it becomes indifferent about action taking. In other words, the agent learns that action taking becomes pointless and thus gives up on it. This can be described as a state of hopelessness (Stephan et al., 2016b).

From the results we conclude that model 2 makes use of its meta-cognitive capabilities to overcome stressful situations in an unhealthy way. In that situation, the agent prefers compensatory over homeostatic control reward. The agent shows decreased reward sensitivity for the homeostatic control outcomes which is a similar effect that we observe in model 1. Adhering to compensatory reward has severe consequences for the agent's understanding of its primary objective, the control problem. The agent, in its helpless state, can be thought of being uninterested in homeostatic control reward for the actual task. This could be interpreted as diminished reward sensitivity in depression. Hopelessness and diminished drive to pursue (pleasurable) activities are symptoms of depression. Model 2 can therefore capture key principles of dyshomeostasis-induced fatigue and depression proposed by Stephan et al. (2016b).

Because of lack of this reward, the agent looses its focus of the control problem. This means that the knowledge about the current state of the continuous dynamical system worsens. This leads the agent from meta-cognitive beliefs about being subjectively helpless to actually having deficits in the other cognitive area. This may be interpreted as objective, potentially diagnosable helplessness. This deficit even lasts beyond the point when the agent recovers. We observe that the agent then needs time to restructure its secondary cognitive factor again to optimally control its bodily state.

It is also interesting to observe our agent's behavior solely during one single, strong perturbation. After a large perturbation, the agent's bodily state is too far away to be able to recover to the set-point in one policy evaluation. It is important to notice that the agent predicts this fact correctly. The key is that the agent chooses a wrong strategy based on subjective optimality, which is a result of the agent configuration and its current beliefs. A psychological disease does not imply that the affected person is unable to act optimal with respect to its beliefs, but the beliefs may be inherently wrong, which is what we observe here. This fact may only be detected by an external observer.

If the agent is not in its helpless state, then it is able to infer the state of the noisy continuous dynamical system. By choosing appropriate actions, the agent offsets and regulates its own noisy actions as well as moderate external perturbations. It accomplishes this with reliance even though the signal is only updated from the underlying continuous dynamics after every episode. It shows that the agent is able to handle the non trivial control problem of taking the most precise action, when left alone.

## VI. Limitations and Future Work

The aim of this work is to provide a purely computational framework that may be used in simulations of experimental disease models influencing allostatic processes. We were able to show in simple examples that this is possible.

During our work we quickly reached the computational limits of the AIF when applied to more complex scenarios. Because of this we had to constrain the size of the parameter space as well as the policy length. We recognize that our model has some drawbacks. We observe that it forgets too quick. This makes it hard to compare it to allostatic processes observed in humans. We assume that this is a result of carrying over the posterior of the bodily state from one episode to the next in both models. With this mechanism the agent forgets any information after a while. Therefore, it is questionable how much the agent learns.

The use of the additional continuous side-car model mostly over-complicated model 2 instead of helping to understand allostatic control better. We see that having the side-car model allows to hide certain aspects from the observing agent but we couldn't make much use of this fact. It might be explored in future work.

Especially in model 2 we observed that the adaption of the meta-cognitive factor was too quick and occurred instantaneously and almost deterministically, even though it is a probabilistic framework. An in-depth analysis of this effect is needed.

In our model, the agent is unable to refuse taking an action. Therefore, the closest to not doing anything is to not care about actions. We cannot model depressive behaviors where the agent just stops doing anything explicitly.

Lastly, we have never made use of the fact that within the AIF, the possibility to learn actions is given. (Cullen et al., 2018) use this feature to navigate the DOOM environment. In our case, all actions of model 1 and 2 are hand-crafted, giving strong prior knowledge on how the agent should act.

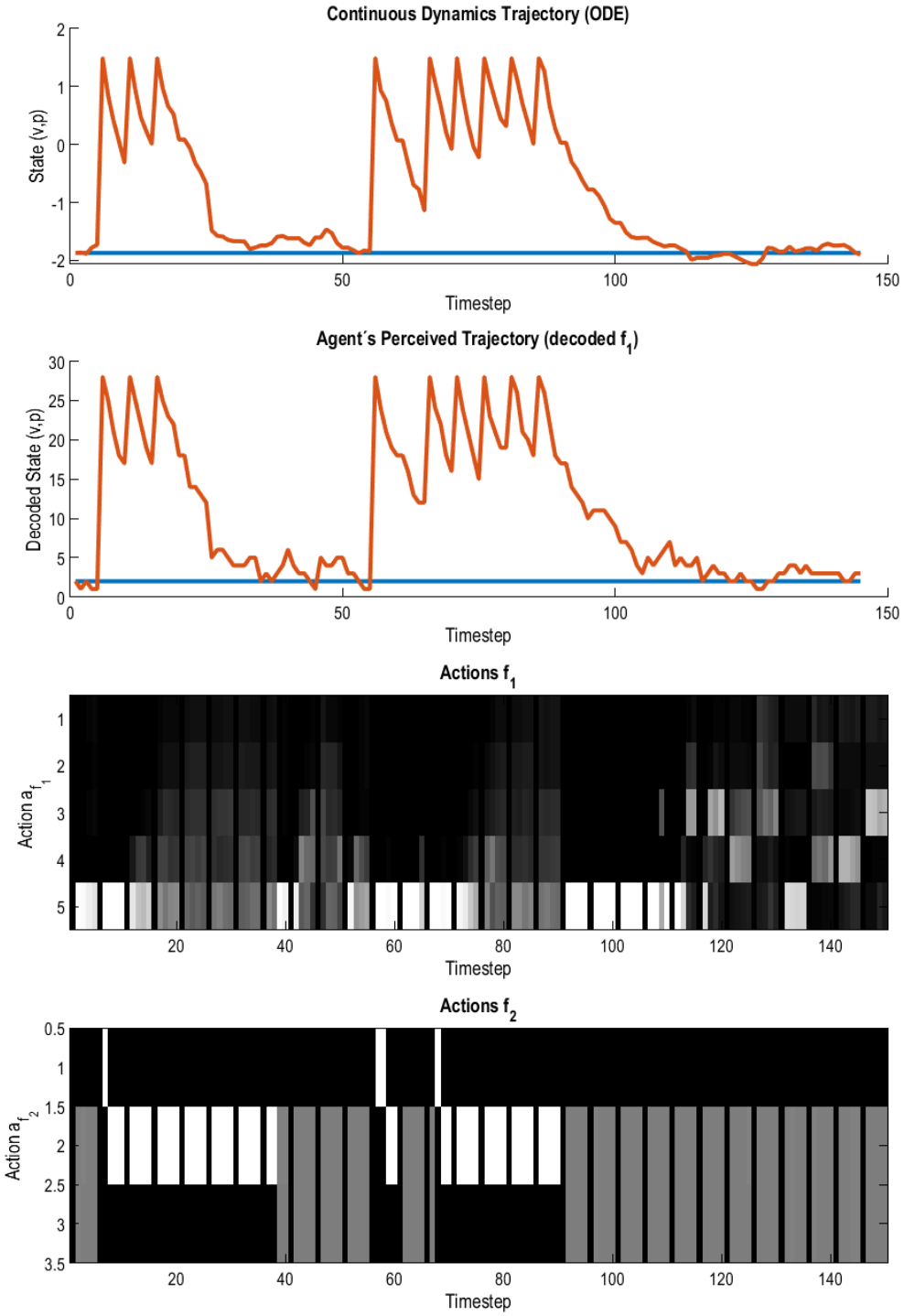


**Fig. 11: True Trajectory**: The set-point $p$ (blue) and the value $v$ (orange) of the continuous dynamical system. **Perceived Trajectory**: $\tilde{p}$ and $\tilde{v}$ decoded from the true state $s^{f_1}$ of the agent's factor $f_1$. **Actions $\mathbf{f_1}$**: Posterior distribution over actions $\tilde{A}^{f_1}$ for the undirected control problem. The distribution is over $K = \kappa_1, \ldots, \kappa_5$ where $\kappa_1$ has the smallest magnitude. **Actions $\mathbf{f_2}$**: Posterior distribution over actions for the meta-cognitive factor. ($1 = a_{a \to h}$, $2 = a_{h \to a}$, $3 = a_s$) The x-axis represents the trial time. Artificial gaps are visible in the action distributions because each episode consists of $T - 1$ actions for $T$ states. This plot shows the same experiment as figure 12.

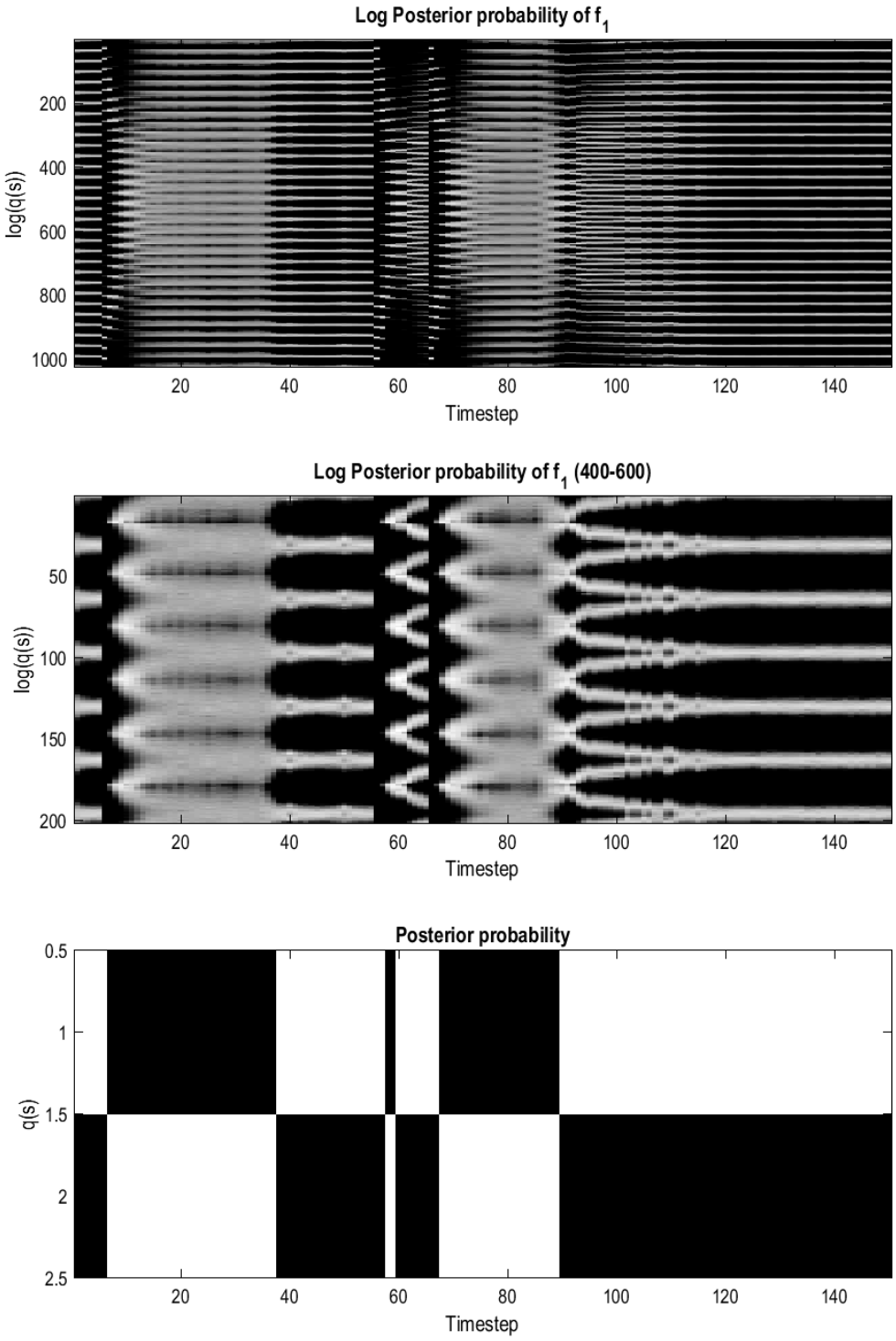


**Fig. 12:** The Bayesian model-averaged posterior $X$ representing $log(q(s))$ over all 150 time steps for factor $f_1$ (**top**) and $f_2$ (**bottom**). $f_1$ shows $(\tilde{v}, \tilde{p})$ encoded as $S_{f_1}$. $f_2$ is the meta-cognitive factor ($1 = s_c$, $2 = s_h$). **center**: Shows a magnification of $s^{f_1}$ for states in the range $[400, 600]$. The plots show the "tailedness" of $q(s^{f_1})$ in confident phases towards the most certain states if the agent is not helpless ($s^{f_2} = s_c$), In contrast to that,we observe how the approximate posterior belief $q(s^{f_1})$ flattens out (becomes more and more uniform), if the agent undergoes the transition to meta-cognitive helplessness ($s^{f_2} = s_h$) (for $t$ in range $[10, 40]$ and $[70, 90]$). The agent is unable to infer on the actual state $f_1$.

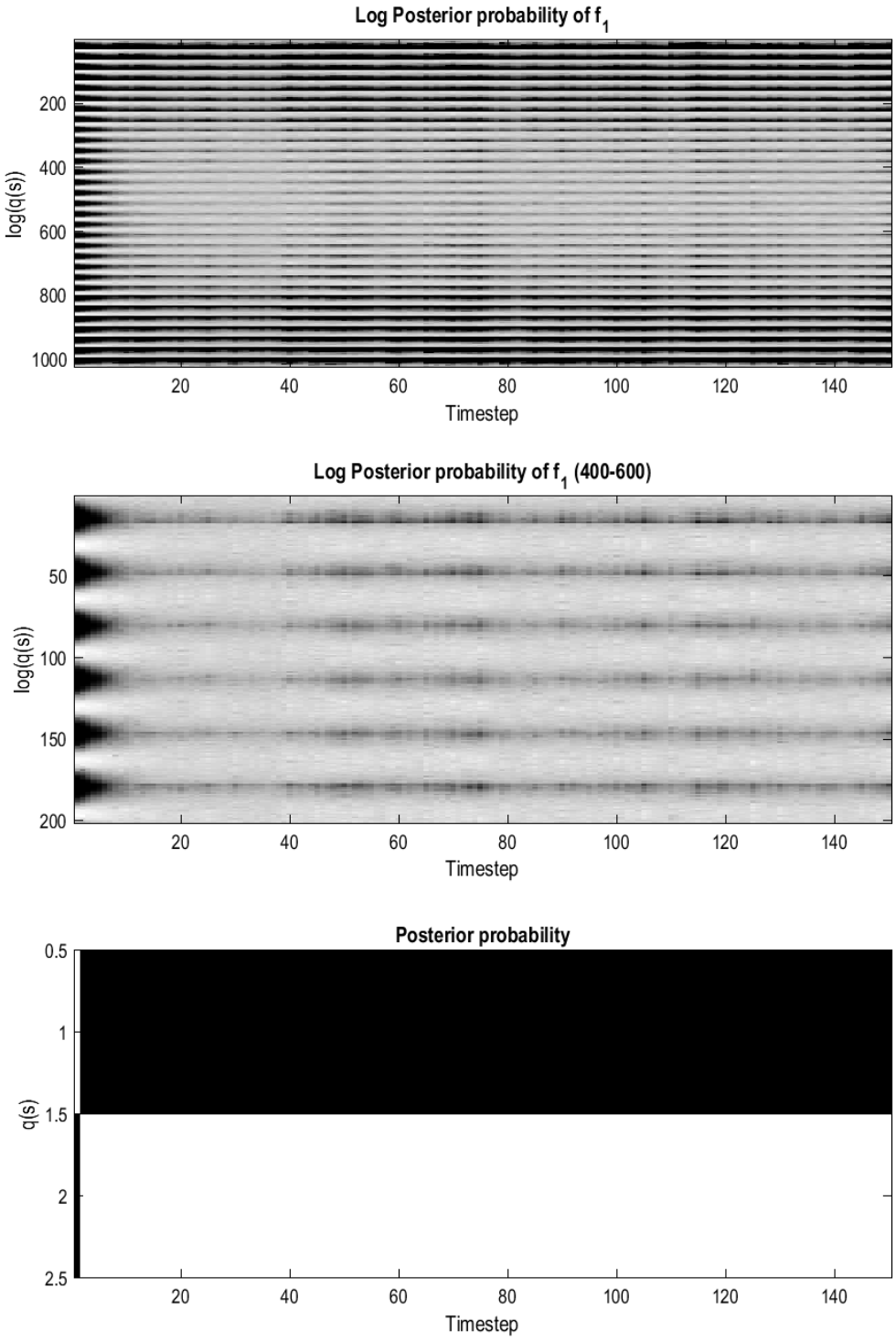


**Fig. 13:** The Bayesian model-averaged posterior $X$ representing $log(q(s))$ over all 150 time steps for factor $f_1$ (**top**) and $f_2$ (**bottom**). $f_1$ shows $(\tilde{v}, \tilde{p})$ encoded as $S_{f_1}$. $f_2$ is the meta-cognitive factor ($1 = s_c$, $2 = s_h$). **center**: Shows a magnification of $s^{f_1}$ for states in the range $[400, 600]$. The agent infers to be in a helpless state ($s^{f_2} = s_c$) because we artificially increased the overall preference of outcome $O_1$'s pseudo state. This has no expressive power and was simply done to ensure the agent stays in a helpless state. We observe no preference for $f_1$ states throughout the entire trial while the agent undergoes the same contextual changes (strong perturbations) as in figure 11.

## REFERENCES


Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym, 2016.

Maell Cullen, Ben Davey, Karl Friston, and Rosalyn Moran. Active inference in openai gym: A paradigm for computational investigations into psychiatric illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3, 07 2018. doi: 10.1016/j.bpsc.2018.06.010.

Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews. Neuroscience*, 11(2):127–138, 2010.

Karl Friston, Jean Daunizeau, and Stefan Kiebel. Reinforcement learning or active inference? *PloS one*, 4:e6421, 02 2009a. doi: 10.1371/journal.pone.0006421.

Karl J. Friston, Jean Daunizeau, and Stefan J. Kiebel. Reinforcement learning or active inference? *PloS one*, 4(7): e6421, 2009b.

Frederike H. Petzschner, Lilian A. E. Weber, Tim Gard, and Klaas E. Stephan. Computational psychosomatics and computational psychiatry: Toward a joint framework for differential diagnosis. *Biological psychiatry*, 82(6):421–430, 2017.

Noor Sajid, Philip J. Ball, and Karl J. Friston. Active inference: demystified and compared, 2019.

Oleg Solopchuk. Tutorial on active inference. URL https://medium.com/@solopchuk/tutorial-on-active-inference-30edcf50f5dc.

Klaas E. Stephan, Dominik R. Bach, Paul C. Fletcher, Jonathan Flint, Michael J. Frank, Karl J. Friston, Andreas Heinz, Quentin J. M. Huys, Michael J. Owen, Elisabeth B. Binder, Peter Dayan, Eve C. Johnstone, Andreas Meyer-Lindenberg, P. Read Montague, Ulrich Schnyder, Xiao-Jing Wang, and Michael Breakspear. Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *The Lancet Psychiatry*, 3(1):77–83, 2016a.

Klaas E. Stephan, Zina M. Manjaly, Christoph D. Mathys, Lilian A. E. Weber, Saee Paliwal, Tim Gard, Marc Tittgemeyer, Stephen M. Fleming, Helene Haker, Anil K. Seth, and Frederike H. Petzschner. Allostatic self-efficacy: A metacognitive theory of dyshomeostasis-induced fatigue and depression. *Frontiers in human neuroscience*, 10:550, 2016b.

Daniel Vigo, Graham Thornicroft, and Rifat Atun. Estimating the true global burden of mental illness. *The Lancet Psychiatry*, 3(2):171–178, 2016.

H. U. Wittchen, F. Jacobi, J. Rehm, A. Gustavsson, M. Svensson, B. Jönsson, J. Olesen, C. Allgulander, J. Alonso, C. Faravelli, L. Fratiglioni, P. Jennum, R. Lieb, A. Maercker, J. van Os, M. Preisig, L. Salvador-Carulla, R. Simon, and H-C Steinhausen. The size and burden of mental disorders and other disorders of the brain in europe 2010. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, 21(9):655–679, 2011.

Ozan Çatal, Tim Verbelen, Johannes Nauta, Cedric De Boom, and Bart Dhoedt. Learning perception and planning with deep active inference, 2020.


## APPENDIX

### A. *Author Contributions*

The following table shows an overview of the contribution of each author to each chapter:

| Chapter | R. Th. | K. Ok.. | O. Kä. |
|---|---|---|---|
| Introduction | | | x |
| Active Inference | x | x | x |
| Modeling of Homeostasis [...] | x | x | x |
| Experiments and Results | x | x | x |
| Discussion | x | x | x |
| Limitations | x | | |
| Coding (Model 1) | | x | |
| Coding (Model 2) | x | | |

### B. *Code*

The project code can be found here:

https://github.com/ratheile/tnu20-active-inference

### C. *Glossary:*

The following abbreviations are used in this paper:

AIF = Active Inference Framework
ELBO = Evidence Lower Bound
OED = Ordinary differential equation
POMDP = Partially Observable Markov Decision Processes
UCP = Undirected Control Problem