



Fachbereich Humanwissenschaften  
Institute of Cognitive Science

Bachelor's Program Cognitive Science

*Exposé*

## **Towards a minimalistic free energy agent**

Esther Chevalier

Matriculation number 972437

10/2020

Esther Chevalier  
Springmannskamp 3  
49090 Osnabrück

E-mail: [echevalier@uos.de](mailto:echevalier@uos.de)  
Tel: 01573 57 20 346

# Contents

<b>1</b>	<b>Exposé</b>	<b>1</b>
1.1	Current state of research . . . . .	1
1.2	Problem . . . . .	1
1.3	Research question . . . . .	2
1.4	Thesis goal and objectives . . . . .	2
1.5	Methods . . . . .	2
1.6	Preliminary thesis structure . . . . .	2
1.7	Preliminary time schedule . . . . .	2
1.8	Relevant literature . . . . .	2

# 1 Exposé

## 1.1 Current state of research

Karl Friston proposes and defends a new formulation habit forming, learning and behavior optimisation in biological agents [1]. The principle he developed over the years rests upon a theoretical framework previously used in statistical thermodynamics. In particular, the concept developed by Richard Feynman, the variational free energy is central to Friston's hypothesis about behaviour learning. Friston has already the free energy framework to formulate and model research questions coming from multiple disciplines, such as computational psychiatry, theoretical biology and artificial intelligence [2]. Friston hypothesize the free energy principle as being more than a simple model of an agent's behavior - it is rather a single feature determining if a biological system is agentic or not.

The free energy principle has a range of properties making it to an elegant model of an agent's behavior and learning patterns. In particular, learning processes in an artificial agent could be described without using external rewards and penalties, as reinforcement learning does. The free energy principle relies on a functional about beliefs about the environmental states the agent can find itself in. The behavior of the agent can be described as "Act to see what you expect to see.". The optimization of a single value, the free energy of each action taken by the agent, leads the agent to make assumptions about the environment and to act optimally given its assumptions and its perceptions. The emerging behavior is characterized by the inferences made by the agent and driven by the action it takes - it is therefore called active inference.

## 1.2 Problem

The free energy principle stays obscure and difficult to understand [3], presumably because of the number of different topics the author addresses. The topic remains primarily investigated by Friston himself, as first or second author. There is a lack of a comprehensive review for early-career scientists who do not have a formal training

in physics or other disciplines necessary to understand the free energy framework. The difficulty to understand the topic is worsen by the unavailability of source code used in free energy agent simulation in Friston’s work e.g. FristonDOOM.

In addition, the free energy principle roots in a deeply physical framework, visible by the vocabulary and concepts used for its formulation []. As stated before, this specialisation hinders the utilisation of the framework by other scientists trained in other fields. The formulation by Friston in his paper in 2006 does not use the vocabulary with the rigour needed. The free energy has since then been adapted to an information theoretic framework [Friston2016]. This reformulation begs the question which concepts are essential for a formulation of the free energy principle and which are not.

### **1.3 Research question**

The goal of this thesis is to implement a free energy agent capable of making inferences in a given environment and learning new behavior based on those inferences. The agent should be as minimalistic as possible to understand what features an agent needs in order to optimize its free energy and its behavior.

### **1.4 Thesis goal and objectives**

### **1.5 Methods**

### **1.6 Preliminary thesis structure**

### **1.7 Preliminary time schedule**

### **1.8 Relevant literature**