

1 Overview

Chance favors the prepared mind.

—Louis Pasteur

1.1 Introduction

This chapter introduces the main question that Active Inference seeks to address: How do living organisms persist while engaging in adaptive exchanges with their environment? We discuss the motivation for addressing this question from a normative perspective, which starts from first principles and then unpacks their cognitive and biological implications. Furthermore, this chapter briefly introduces the structure of the book, including its subdivision into two parts: the first of which aims to help readers understand Active Inference, and the second of which aims to help them use it in their own research.

1.2 How Do Living Organisms Persist and Act Adaptively?

Living organisms constantly engage in reciprocal interactions with their environment (including other organisms). They emit *actions* that change the environment and receive *sensory observations* from it, as schematically illustrated in figure 1.1.

Living organisms can only maintain their bodily integrity by exerting adaptive *control* over the action-perception loop. This means acting to solicit sensory observations that either correspond to desired outcomes or goals (e.g., the sensations that accompany secure nutrients and shelter for simple

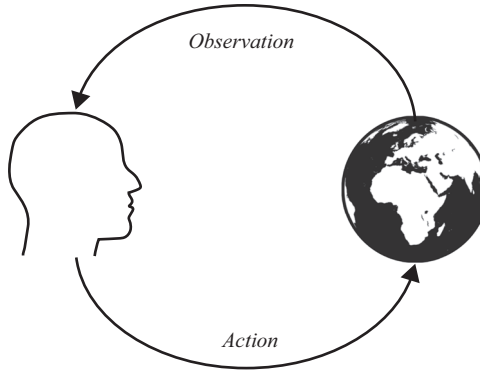


Figure 1.1

An action-perception cycle reciprocally connecting a creature and its environment. The term *environment* is intentionally generic. In the examples that we discuss, it can include the physical world, the body, the social environment, and so on.

organisms, or friends and jobs for more complex ones) or help in making sense of the world (e.g., informing the organism about its surroundings).

Engaging in adaptive action-perception loops with the environment poses formidable challenges to living organisms. This is largely due to the recursive nature of the cycle, where each observation, solicited by the previous action, changes how we decide on the next action, to solicit the next observation. The possibilities for control and adaptation are plentiful, but very few are useful. Yet during evolution, living organisms have managed to develop adaptive strategies to face the fundamental challenges of existence. These strategies vary in their level of cognitive sophistication, with simpler and more rigid solutions in simpler organisms (e.g., following nutrient gradients in bacteria) and more cognitively demanding and flexible solutions in more advanced organisms (e.g., planning to achieve distal goals in humans). These strategies also vary for the timescales at which they are selected and operate—ranging from simple responses to environmental threats or morphological adaptations that arise at an evolutionarily timescale, to behavioral patterns established during cultural or developmental learning, up to those requiring cognitive processes that operate at comparable timescales to action and perception (e.g., attention and memory).

1.3 Active Inference: Behavior from First Principles

This diversity is a blessing for biology but challenging for formal theories of brain and mind. Broadly, there are two perspectives we could take on this. One perspective is that different biological adaptations, neural processes (e.g., synaptic exchanges and brain networks), and cognitive mechanisms (e.g., perception, attention, social interaction) are highly idiosyncratic and require dedicated explanations. This would lead to proliferation of theories in fields like philosophy, psychology, neuroscience, ethology, biology, artificial intelligence, and robotics, with little hope for their unification. Another perspective is that, despite their diverse manifestations, the central aspects of behavior, cognition, and adaptation in living organisms are amenable to a coherent explanation from first principles.

These two possibilities map to two different research programs and, to some extent, different attitudes toward science: “neats” versus “scruffies” (terms due to Roger Shank). Neats always seek unification beyond the (apparent) heterogeneity of brain and mind phenomena. This usually corresponds to designing top-down, normative¹ models that start from first principles and try to derive as much as possible about brains and minds. Scruffies instead embrace the heterogeneity by focusing on details that demand dedicated explanations. This usually corresponds to designing bottom-up models that start from data and use whatever works to explain complex phenomena, including different explanations for different phenomena.

Is it possible to explain heterogeneous biological and cognitive phenomena from first principles, as the neats assume? Is a unified framework to understand brain and mind possible?

This book answers these questions affirmatively and advances Active Inference as a normative approach to understand brain and mind. Our treatment of Active Inference starts from first principles and unpacks their cognitive and biological implications.

1.4 Structure of the Book

The book comprises two parts. These are aimed at readers who want to understand Active Inference (first part) and those who seek to use it for their own research (second part). The first part of the book introduces Active Inference both conceptually and formally, contextualizing it within current theories of cognition. The goal of this first part is to provide a comprehensive, formal,

and self-contained introduction to Active Inference: its main constructs and implications for the study of brain and cognition.

The second part of the book illustrates specific examples of computational models that use Active Inference to explain cognitive phenomena, such as perception, attention, memory, and planning. The goal of this second part is to help readers both understand existing computational models using Active Inference and design novel ones. In short, this book divides into theory (part 1) and practice (part 2).

1.4.1 Part 1: Active Inference in Theory

Active Inference is a normative framework to characterize Bayes-optimal² behavior and cognition in living organisms. Its normative character is evinced in the idea that all facets of behavior and cognition in living organisms follow a unique imperative: *minimizing the surprise of their sensory observations*. *Surprise* has to be interpreted in a technical sense: it measures how much an agent's current sensory observations differ from its preferred sensory observations—that is, those that preserve its integrity (e.g., for a fish, being in the water). Importantly, minimizing surprise is not something that can be done by passively observing the environment: rather, agents must adaptively *control* their action-perception loops to solicit desired sensory observations. This is the active bit of Active Inference.

Minimizing surprise turns out to be a challenging problem for technical reasons that will become apparent later. Active Inference offers a solution to this problem. It assumes that even if living organisms cannot directly minimize their surprise, they can minimize a proxy—called (*variational*) *free energy*. This quantity can be minimized through neural computation in response to (and in anticipation of) sensory observations. This emphasis on free energy minimization discloses the relation between Active Inference and the (first) principle that motivates it: the *free energy principle* (Friston 2009).

Free energy minimization seems a very abstract starting point to explain biological phenomena. However, it is possible to derive a number of formal and empirical implications from it and to address a number of central questions in cognitive and neural theory. These include how the variables involved in free energy minimization may be encoded in neuronal populations; how the computations of minimized free energy map to specific cognitive processes, such as perception, action selection, and learning; and

what kind of behaviors emerge when an Active Inference agent minimizes its free energy.

As the above list of topics exemplifies, in this book we are mainly concerned with Active Inference and free energy minimization at the level of living organisms—simpler (e.g., bacterial) or more complex (e.g., human)—and their behavioral, cognitive, social, and neural processes. This clarification is necessary to contextualize our treatment of Active Inference within the more general free energy principle (FEP), which discusses free energy minimization across a much wider range of biological phenomena and timescales beyond neural information processing—ranging from evolutionary to cellular and cultural (Friston, Levin et al. 2015; Isomura and Friston 2018; Palacios, Razi et al. 2020; Veissière et al. 2020)—which are beyond the scope of this book.

It is possible to motivate Active Inference by taking one of two roads: a high road and a low road; see figure 1.2. These two roads provide two distinct but highly complementary perspectives on Active Inference:

- The *high road* to Active Inference starts from the question of how living organisms persist and act adaptively in the world and motivates

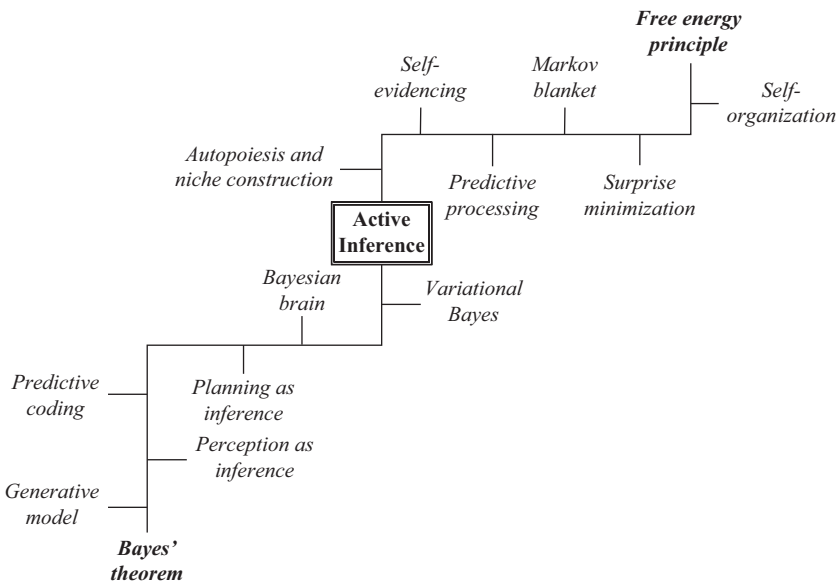


Figure 1.2
Two roads to Active Inference: the high road (starting from top-right) and the low road (starting from bottom-left).

Active Inference as a normative solution to these problems. This high road perspective is useful to understand the normative nature of Active Inference: *what* living organisms must do to face their fundamental existential challenges (minimize their free energy) and *why* (to vicariously minimize the surprise of their sensory observations).

- The *low road* to Active Inference starts from the notion of the Bayesian brain, which casts the brain as an inference engine trying to optimize probabilistic representations of the causes of its sensory input. It then motivates Active Inference as a specific, variational approximation to the (otherwise intractable) inferential problem, which has a degree of biological plausibility. This low road perspective is useful to illustrate *how* Active Inference agents minimize their free energy—therefore illustrating Active Inference not just as a principle but also as a mechanistic explanation (aka *process theory*) of cognitive functions and their neuronal underpinnings.

In chapter 2, we set out the low road perspective on Active Inference. We start from foundational theories that cast perception as a problem of statistical (Bayesian) inference (Helmholtz 1866) and their modern incarnation in the Bayesian brain hypothesis (Doya 2007). We will see that to perform such (perceptual) inference, living organisms must be equipped with—or embody—a *probabilistic generative model* of how their sensory observations are generated, which encodes beliefs (probability distributions) about both observable variables (sensory observations) and nonobservable (hidden) variables. We will extend this inferential view beyond perception to cover problems of action selection, planning, and learning.

In chapter 3, we will illustrate the complementary high road perspective on Active Inference. This chapter introduces the FEP and the imperative for biological organisms to minimize surprise. Further to this, it unpacks how this principle encompasses the dynamics of self-organization and the preservation of a statistical boundary or *Markov blanket* that maintains separation from the environment. This is vital in maintaining the integrity of biological creatures, and it is central to their autopoiesis.

In chapter 4, we will unpack Active Inference more formally. This chapter takes its cue from the discussion of the Bayesian brain in chapter 2 and sets out the mathematical relationship between the self-evidencing dynamics of chapter 3 and variational inference. In addition, this chapter sets out two

sorts of generative model used to formulate Active Inference problems. These include the partially observed Markov decision processes used for decision-making and planning and the continuous time dynamical models that interface with sensory receptors and muscles. Finally, we see how free energy minimization for each of these models manifests as dynamic belief updating.

In chapter 5, we will move from formal treatments to biological implications of Active Inference. By starting from the premise that “everything that changes in the brain must minimize free energy” (Friston 2009), we will discuss how the specific quantities involved in the free energy minimization (e.g., prediction, prediction error, and precision signals) manifest in neuronal dynamics. This aids in mapping the abstract computational principles of Active Inference to specific neural computations that can be executed by physiological substrates. This is important in forming hypotheses under this framework and ensures that these are answerable to measured data. In other words, chapter 5 sets out the process theory associated with Active Inference.

Throughout the first part of the book, we will discuss several characteristic aspects of Active Inference. These highlight the ways in which it is different from alternative frameworks that seek to explain biological regulation and cognition—some of which we preview here.

- Under Active Inference, *perception* and *action* are two complementary ways to fulfill the same imperative: minimization of free energy. Perception minimizes free energy (and surprise) by (Bayesian) belief updating or *changing your mind*, thus making your beliefs compatible with sensory observations. Instead, action minimizes free energy (and surprise) by *changing the world* to make it more compatible with your beliefs and goals. This unification of cognitive functions marks a fundamental difference between Active Inference and other approaches that treat action and perception in isolation from one another. Learning is yet another way to minimize free energy. However, it is not fundamentally different from perception; it simply operates at a slower timescale. The complementarity between perception and action will be unpacked in chapter 2.
- In addition to driving action selection in the present to change currently available sensory data, the Active Inference framework accommodates planning—or the selection of the optimal course of action (or policy) in the future. Optimality here is measured in relation to an *expected free*

energy and is distinct from the notion of *variational free energy* considered above in the context of action and perception. Indeed, while computing variational free energy depends on present and past observations, computing expected free energy also requires predicted future observations (hence the term *expected*). Interestingly, the expected free energy of a policy comprises two parts. The first quantifies the extent to which the policy is expected to resolve uncertainty (exploration) and the second how consistent the predicted outcomes are with an agent's goals (exploitation). In contrast with other frameworks, policy selection in Active Inference automatically balances exploration and exploitation. The relations between variational and expected free energy will be unpacked in chapter 2.

- Under Active Inference, all cognitive operations are conceptualized as inference over *generative models*—in keeping with the idea that the brain performs probabilistic computations—aka *the Bayesian brain hypothesis*. Yet, the appeal to a specific approximate form of Bayesian inference—that is, a variational scheme that is motivated by first principles—adds specificity to the process theory. Furthermore, Active Inference extends the inferential approach to domains of cognition that are rarely considered and adds some specificity to the kind of models and inferential processes that may be implemented by biological brains. Under some assumptions, the dynamics that emerge from generative models used in Active Inference closely correspond to widespread models in computational neuroscience, such as predictive coding (Rao and Ballard 1999) and the Helmholtz machine (Dayan et al. 1995). The specifics of the variational scheme will be unpacked in chapter 4.
- Under Active Inference, both perception and learning are *active* processes, for two reasons. First, the brain is essentially a *predictive machine*, which constantly predicts incoming stimuli rather than passively waiting for them. This is important as perceptual and learning processes are always contextualized by prior predictions (e.g., expected and unexpected stimuli affect perception and learning in different ways). Second, creatures engaging in Active Inference actively seek out *salient* sensory observations that resolve their uncertainty (e.g., by orienting their sensors or selecting learning episodes that are informative). The active character of perception and learning stands in contrast with most current theories that treat them as largely passive processes; this will be unpacked in chapter 2.

- Action is quintessentially goal directed and purposive. It starts from a desired outcome or goal (analogous to the concept of a set-point in cybernetics), which is encoded as a prior prediction. Planning proceeds by inferring an action sequence that fulfills this prediction (or equivalently, reduces any prediction error between prior prediction and the current state). The goal-directed character of action in Active Inference is in keeping with early cybernetic formulations but is distinct from most current theories that explain behavior in terms of stimulus-response mappings or state-action policies. Stimulus-response or habitual behavior then becomes a special case of a broader family of policies in Active Inference. The goal-directed nature of Active Inference will be unpacked in chapters 2 and 3.
- Various constructs of Active Inference have plausible biological analogues in the brain. This implies that—once one has defined a specific generative model for a problem at hand—one can move from Active Inference as a normative theory to Active Inference as a process theory, which makes specific empirical predictions. For example, perceptual inference and learning correspond to changing synaptic activity and changing synaptic efficacy, respectively. Precision of predictions (in predictive coding) corresponds to the synaptic gain of prediction error units. Precision of policies corresponds to dopaminergic activity. Some of the biological consequences of Active Inference will be unpacked in chapter 5.

1.4.2 Part 2: Active Inference in Practice

While the first part of the book provides readers with the conceptual and formal tools to understand Active Inference, the second part focuses on practical issues. Specifically, we hope to provide readers with the tools to understand existing Active Inference models of cognitive functions (and dysfunctions) and to design novel ones. To this aim, we discuss specific examples of models using Active Inference. Importantly, models of Active Inference can vary along different dimensions (e.g., with discrete or continuous time formulations, flat or hierarchical inference). The second part is structured as follows:

In chapter 6, we introduce a recipe to build Active Inference models. The recipe covers the essential steps to design an effective model, which include the identification of the system of interest, the most appropriate

form of the generative model (e.g., to characterize discrete- or continuous-time phenomena), and the specific variables to be included in the model. This chapter therefore offers an introduction to the design principles that underwrite the models discussed in the following chapters.

In chapter 7, we discuss Active Inference models that address problems formulated in discrete time; for example, as hidden Markov models (HMMs) or partially observable Markov decision processes (POMDPs). Our examples include a model of perceptual processing and a model of discrete foraging choices—that is, whether to turn left or right at a decision point to secure a reward. We also introduce topics such as information seeking, learning, and novelty seeking, which can be treated in terms of discrete-time Active Inference.

In chapter 8, we discuss Active Inference models that address problems formulated in continuous time, using stochastic differential equations. These include models of perception (like predictive coding), movement control, and sequential dynamics. Interestingly, it is in the continuous-time formulation that some of the most distinctive predictions of Active Inference appear, such as the idea that movement generation stems from the fulfillment of predictions and that attentional phenomena can be understood in terms of precision control. We also introduce hybrid models of Active Inference that include both discrete- and continuous-time variables. These permit simultaneous assessment of the choice among discrete options (e.g., targets for saccades) and the continuous movements resulting from the choice (e.g., oculomotor movements).

In chapter 9, we illustrate how to use Active Inference models to analyze data from behavioral experiments. We discuss the specific steps that are necessary for model-based data analysis, from the collection of data to the formulation of a model and its inversion to support the analysis of data from single participants or at the group level.

In chapter 10, we discuss the relations between Active Inference and other theories in psychology, neuroscience, AI, and philosophy. We also highlight the most important aspects of Active Inference that distinguish it from the other theories.

In the appendixes, we briefly discuss the mathematical background required to understand the most technical parts of the book, including the notions of Taylor series approximation, variational Laplace, variational

calculus, and more. For reference we also present in a concise form the most important equations used in Active Inference.

In sum, the second part of the book illustrates a broad variety of models of biological and cognitive phenomena that can be constructed using Active Inference and a methodology to design novel ones. Apart from the interest of the specific models, we hope that our treatment clarifies the value of using a unified, normative framework to address biological and cognitive phenomena from a coherent perspective. In the end, this is the real appeal of normative frameworks: to provide a unified perspective and a guiding principle to reconcile apparently disconnected phenomena—in this case, phenomena like perception, decision-making, attention, learning, and movement control, each having its separate chapter in any psychology or neuroscience manual.

The models highlighted in the second part have been selected to illustrate specific points as simply as possible. While we cover several models and domains, from discrete-time decisions to continuous-time perception and movement control, we are clearly disregarding many others that are equally interesting. Many other Active Inference models exist in the literature that cover domains as diverse as biological self-organization and the origins of life (Friston 2013), morphogenesis (Friston, Levin et al. 2015), cognitive robotics (Pio-Lopez et al. 2016, Sancaktar et al. 2020), social dynamics and niche construction (Bruineberg, Rietveld et al. 2018), the dynamics of synaptic networks (Palacios, Isomura et al. 2019), learning in biological networks (Friston and Herreros 2016), and psychopathological conditions, such as post-traumatic stress disorder (Linson et al. 2020) and panic disorder (Maisto, Barca et al. 2021). These models vary along many dimensions: some are more directly related to biology whereas others are less so; some are single-agent models whereas others are multi-agent models; some target adaptive inference whereas other target maladaptive inference (e.g., in patient groups), and so on.

This growing literature exemplifies the increasing popularity of Active Inference and the possibility of using it in a very large variety of domains. The aim of this book is to provide our readers with the ability to understand and use Active inference in their own research—possibly, to explore its unforeseen potentialities.

1.5 Summary

This chapter briefly introduces the Active Inference approach to explain biological problems from a normative perspective—and previews some implications of this perspective that will be unpacked in later chapters. Furthermore, this chapter highlights the division of the book into two parts, which aim to help readers understand Active Inference and use it in their own research, respectively. Over the next few chapters, we will develop the low road and high road perspectives outlined herein, before delving into the structure of generative models and the resulting message passing. Together these comprise Active Inference in principle and provide the preliminaries for Active Inference in practice. We hope that these chapters will persuade readers that Active Inference offers not only a unifying principle under which to understand behavior but also a tractable approach to studying action and perception in autonomous systems.