Biological Psychology 186 (2024) 108741

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Review

Active inference as a theory of sentient behavior

Giovanni Pezzulo [a,*], Thomas Parr [b], Karl Friston [c,d]

[a] Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy
[b] Nuffield Department of Clinical Neurosciences, University of Oxford, UK
[c] Wellcome Centre for Human Neuroimaging, Queen Square Institute of Neurology, University College London, London, UK
[d] VERSES AI Research Lab, Los Angeles, CA 90016, USA

ARTICLE INFO

Keywords:
Active inference
Predictive coding
Generative model

ABSTRACT

This review paper offers an overview of the history and future of active inference—a unifying perspective on action and perception. Active inference is based upon the idea that sentient behavior depends upon our brains' implicit use of internal models to predict, infer, and direct action. Our focus is upon the conceptual roots and development of this theory of (basic) sentience and does not follow a rigid chronological narrative. We trace the evolution from Helmholtzian ideas on unconscious inference, through to a contemporary understanding of action and perception. In doing so, we touch upon related perspectives, the neural underpinnings of active inference, and the opportunities for future development. Key steps in this development include the formulation of predictive coding models and related theories of neuronal message passing, the use of sequential models for planning and policy optimization, and the importance of hierarchical (temporally) deep internal (i.e., generative or world) models. Active inference has been used to account for aspects of anatomy and neurophysiology, to offer theories of psychopathology in terms of aberrant precision control, and to unify extant psychological theories. We anticipate further development in all these areas and note the exciting early work applying active inference beyond neuroscience. This suggests a future not just in biology, but in robotics, machine learning, and artificial intelligence.

1. Introduction

Psychologists and neuroscientists are increasingly entertaining the idea of the brain as a "prediction machine", which learns an internal (i.e., generative) model of the lived world – and of the consequences of its actions – to make sense of sensations, predict how the current situation will unfold (i.e., learning and perception), and to act in a purposeful manner (i.e., action selection, exploration-exploitation, planning, et cetera). This idea appears in several guises, including the *Bayesian brain*, the *predictive brain*, *predictive processing*, *predictive coding*, *active inference* and the *free energy principle*, to name a few.

Here, we critically review the origins, scope and impact of this idea, in fields like psychology and neuroscience. For conceptual clarity, we focus specifically on *active inference*: a normative theory of sentient behavior that formalizes the "predictive brain" idea and provides a first-principle account of its computational and neuronal processes (Parr et al., 2022).

While active inference is still relatively young, it has a growing impact across various disciplines. It is increasingly used by (for example) neuroscientists interested in the neural circuits supporting predictions and prediction errors (Bastos et al., 2012; Parr & Limanowski, Rawji, et al., 2021; Parr & Friston, 2018; Walsh et al., 2020); psychologists interested in how we deal with uncertainty and cognitive effort during decision-making (Parr et al., 2023; Rens et al., 2023), modelers interested in the mechanisms of action-perception, exploration-exploitation and higher cognition (Friston, FitzGerald, et al., 2017; Friston, Lin, et al., 2017; Pezzulo et al., 2015, 2018), clinicians interested in understanding aberrant behavior in psychopathology (Maisto et al., 2021; Van den Bergh et al., 2017), roboticists interested in self-supervised learning of world models and goal-directed behavior (Ahmadi & Tani, 2019; Taniguchi et al., 2023), and neurophilosophers (Clark, 2015; Hohwy, 2013).

This breadth of application is appealing, but risks creating a fragmented picture and some uncertainty about its original commitments and conceptual implications. The aim of this brief manuscript is to help researchers using (or interested in) predictive coding and active inference to "connect the dots" and orient themselves within a growing literature. Despite distinct lines of work — that emphasize different aspects of active inference — these applications all rest on the same core

* Correspondence to: Institute of Cognitive Sciences and technologies, National Research Council, Via S. Martino della Battaglia 44, 00185 Rome, Italy.
E-mail address: giovanni.pezzulo@istc.cnr.it (G. Pezzulo).

https://doi.org/10.1016/j.biopsycho.2023.108741
Received 18 July 2023; Received in revised form 5 December 2023; Accepted 29 December 2023
Available online 3 January 2024

principles. To foreground these core principles, we will look at the historical and conceptual origins of active inference—to illustrate how its core principles were introduced; then consider briefly how the scope of active inference has expanded into several disciplines—and finally look to future developments. Given the brevity of this treatment, we cannot provide a full introduction to active inference. Rather, we provide an overview of the narrative in (Parr et al., 2022), which interested readers can consult.

In the next section, we briefly discuss the conceptual (and historical) roots of active inference in early views of prediction and action-based cognition. We then review some key developments of active inference, by focusing on landmark papers that explain how it stems from a single principle (namely, free energy minimization). We next consider its scope across perception, action, planning, etc. This brief review helps us make the point that active inference provides a unifying perspective on several cognitive topics and theories and across levels of understanding, from conceptual to neural. Finally, we briefly highlight some promising research directions that could expand the scope of active inference – and potentially its impact on psychology and neuroscience.

2. The conceptual and historical roots of active inference

Active inference has roots in various early theories in cognitive science (and beyond, in fields that would not necessarily use the label "cognitive"). One root is the idea that the brain carries a small-scale model of the environment and uses it to mentally simulate *what-if actions*, instead of (or before) acting on the environment (Craik, 1943). This idea is foundational in cognitive science. For example, (Tolman, 1948) proposed that humans, rodents and other animals find their way in a maze by first learning a mental model or "cognitive map", rather than by considering which of their navigation actions were previously rewarded the most, as assumed by behaviorist formulations.

Another root is the idea of (Helmholtz, 1866) that perception is an (unconscious) inference based on an internal generative model – that uses recurrent (top-down and bottom-up) counter-streams of processing, rather than bottom-up transduction of external sensations into internal representations (and later actions). This idea was later developed in psychology (Gregory, 1968, 1980) and computational neuroscience; giving rise to the "Bayesian brain" hypothesis (Doya et al., 2007) and to formulations of predictive coding as a possible neurobiological implementation of perception-as-inference in the brain (Friston, 2005; Rao & Ballard, 1999). Beyond perception, other cognitive functions were later described in terms of inference, i.e., planning-as-inference (Botvinick & Toussaint, 2012).

Yet another "root" is the idea of cyberneticists (Miller et al., 1960; Powers, 1973; Wiener, 1948) that goal-directed action proceeds by firstly setting up a desired state or observation (e.g., feeling warm), then monitoring the discrepancy – now referred to as a "prediction error" – between the preferred and sensed state (e.g., feeling excessively warm), and then selecting a course of action that reduces this discrepancy – where "action" is a suitcase word and can include any means to exert control over external stimuli; ranging from simple autonomic reflexes (e.g. thermoregulation) to sophisticated plans (e.g., visiting one's favorite ice cream shop). A key result in this field – which coheres with the Helmholtzian perspective above – is the 'Good regulator theorem' of (Conant & Ashby, 1970), which argues that effective regulatory systems must [be a] model the environment they regulate. In a similar vein, in psychology, ideomotor theory proposed that action control is essentially anticipatory and that action are selected and controlled by their anticipated consequences or outcomes, not through stimulus-response (Hoffmann, 2003; Hommel, 2003; James, 1890).

Besides cybernetics, there are other influential views that highlight the centrality of adaptive regulation for behavior and life itself. One example is the idea that living organisms are autopoietic systems, which create the conditions for their own existence. More recently, this idea has been framed as 'self-evidencing' (Hohwy 2016) – i.e., creatures seek out sensations that provide evidence for their continued existence. Intuitively, sensing our body temperature to be around 37 °C offers more evidence that we are still alive than body temperatures far from this value. The concept of autopoiesis gave birth to enactive approaches in philosophy (Maturana & Varela, 1980). From another angle, it has been postulated that a central imperative for living organisms is maintenance of physiological homeostasis (i.e., correction of deviations from preferred physiological states through reflexive actions) and the regulation of basic imperatives (Cannon, 1929) – but more modern theories emphasize that physiological regulation is fundamentally anticipatory (i.e., allostatic) (Sterling, 2012). Various researchers have proposed that closed-loop adaptive regulation (and not stimulus-response) is key to understanding not just physiology but (potentially) all cognitive processing (Cisek, 1999; Pezzulo & Cisek, 2016).

Finally, another root is the idea that cognitive processes, such as learning, perception and decision-making, require an active engagement of organisms with the environment. One early example of this action-oriented perspective is the view of Gibson that perceiving things consists in seeing what to do or not to do with them, i.e., perceiving affordances (Gibson, 1979). More recently, various researchers proposed the necessity of a "pragmatic turn" in cognitive science and neuroscience – and the need to recognize the importance of action as part and parcel of our cognition (Buzsaki, 2019; Cisek & Kalaska, 2010; Cisek & Pastor-Bernier, 2014; Engel et al., 2016; Lepora & Pezzulo, 2015; O'Regan & Noe, 2001), rather than just a way to report "central" decisions, as assumed in conventional (serial) theories.

Interestingly, each of these ideas implies a shift from reactive to predictive, enactive views of the brain. While a reactive brain waits for incoming stimuli, a predictive and active brain predicts external events (e.g., predictive coding) and actively gathers evidence (i.e., active sensing and active learning) to make sense of the world. While a reactive brain selects actions based on the past and present (e.g., the history of reinforcement and the current cue), a predictive brain actively imagines its preferred future and then makes this happen by acting (e.g., acts in a goal-directed manner). While a reactive brain maintains homeostasis, a predictive brain acts to anticipate needs and performs anticipatory regulatory (or allostatic) actions.

All these (and other) views contributed to raising the importance of predictive and enactive views of the brain and of cognition. However, each of these perspectives were somewhat disconnected from one another and linked to different research traditions, which are sometimes seen as conflicting with one another (e.g., the Helmholtzian and the Gibsonian traditions). One benefit of active inference is that it helps unify and thereby advance these traditions, as we explain in the following Sections.

3. The normative perspective of active inference – and how it has developed

Active inference provides a normative perspective that unifies and advances the predictive and enactive views of brain and behavior. It does so by highlighting that several apparently disconnected accounts – identified by early theories – stem parsimoniously from the assumption that living organisms obey a single imperative: namely, they act to minimize their *surprise,*[1] or more formally, their *variational free energy*.

The mathematics of variational free energy minimization is beyond the scope of this article; we suggest to the interested readers to consult (Parr et al., 2022). Here, instead, we introduce the key concepts of the theory, by briefly reviewing (non-chronologically) selected landmark papers and linking them to the early theories.

Active inference starts from a simple consideration: that to maintain

[1] Technically, surprise here refers to self-information (a.k.a., surprisal); namely, the implausibility of some (sensory) outcome under a (generative) model of how that outcome was generated.

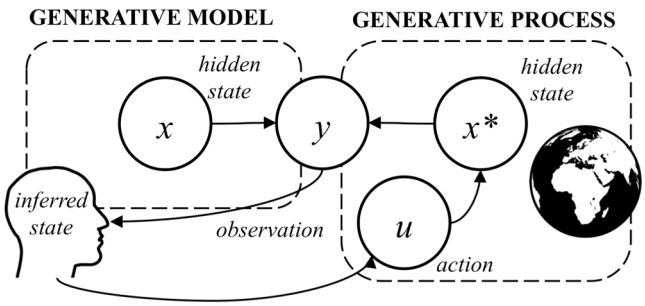

Fig. 1. Generative model and generative process in active inference. This Figure—reproduced from (Parr et al., 2022)—illustrates the structure of active inferential theories of brain function. Our worlds evolve according to some dynamical process that generates observations (y) from hidden states (x *). Our internal models account for observations in terms of hypothetical hidden states (x). Our inferences about these states based upon our observations then drive actions (u) that intervene on the processes generating our sensations.

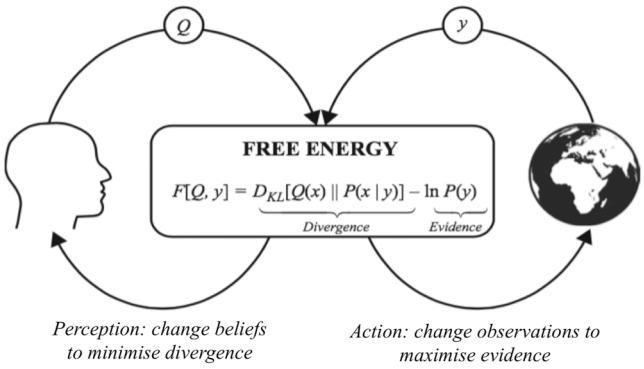

Fig. 2. Perception and action play complementary roles in the minimization of variational free energy. This Figure— reproduced from (Parr et al., 2022)—highlights the relationship between action and perception via free energy (F). Perception involves minimizing free energy by changing our beliefs (Q) about states (x). This effectively minimizes the divergence (D_{KL}) between our beliefs and the probability of these states given sensory data (y). Action minimizes free energy through changing those parts of the free energy that depend upon sensory data—notably, the evidence or probability of data under our internal model.

their existence and integrity, all living organisms need to remain in a bounded set of characteristic states that basically define their place within an ecological niche; for example, a fish cannot live out of water. Using the lexicon of Bayesian inference, being out of water for a fish is a "surprising" state. Clearly a fish should avoid this surprise, and the idea generalizes to suggest that living organisms must avoid surprising states (Friston et al., 2010). If they did not, they would not be living organisms for long. Another way of looking at this is that everything (including me) is defined by being in some characteristic (attracting) set of states. Conversely, I am defined by the kinds of states I cannot be in. These are surprising states.

A computationally tractable solution to surprise minimization is the minimization of an information-theoretic quantity – variational free energy – which is a function of two things: a generative model (i.e., a statistical model that describes how sensations are generated) and observed sensory data. This implies that a living organism must be equipped with a generative model – or in the lexicon of (Craik, 1943), a small-scale model – to predict the sensations generated by the world (and by the organism's place in it). In Bayesian terms, a generative model comprises two things: a *prior* over the hidden (i.e., unobserved) variables of interest and a *likelihood function* that maps the hidden variables to observables (Bishop, 2006). See Fig. 1 for a schematic illustration of the organism's generative model of the world and its relation with the generative process: the true environmental contingencies that generate its observations, which is inaccessible to the organism.

Put simply, an organism can minimize variational free energy by aligning the predictions of its generative model and the data it observes. In different settings, this minimization has been described in various ways, such as the minimization of surprise, of prediction errors, or of the discrepancy between the model and the world. All of these are equivalent to the minimization of variational free energy under specific sets of assumptions.

Interestingly, aligning the predictions derived from a generative model and data can be achieved in two ways: by changing the model *predictions* and by changing the observed *data*. The former corresponds to revising the agent's beliefs (used in the technical sense of probability distributions over hidden variables) if they do not explain the data well. This is exactly the inferential view of perception of (Helmholtz, 1866). The latter corresponds to acting in the world to change the data that will be sampled next – to render them more like the organism's prior predictions. This latter perspective on action – and on its dependence on expected outcomes – is highly congruent with cybernetics (Miller et al., 1960; Powers, 1973; Wiener, 1948) and ideomotor theory (Hoffmann, 2003; Hommel, 2003; James, 1890).

In sum, changing beliefs about the causes of data (i.e., perception) and changing the data (i.e., action) are two aspects of free energy

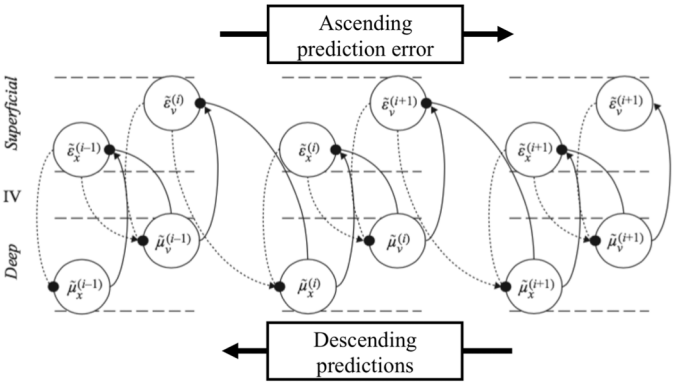

Fig. 3. The architecture of predictive coding. This Figure—reproduced from (Parr et al., 2022)—shows the message passing between populations of neurons under a predictive coding scheme as it might manifest in the layers of the cerebral cortex (separated into superficial layers I-III, layer IV, and deep layers V-VI). This shows predictions based upon expectations (μ) being subtracted from ascending signals to compute errors (ε), which are used to update expectations. The subscripts indicate whether we are dealing with fast changing dynamical variables (x) or more slowly changing contextual variables (v) which act to link together different hierarchical levels, with hierarchy indicated by the bracketed superscripts. As we ascend the hierarchy, the variables we deal with become slower, such that the contextual variables at one hierarchical level evolve over the same timescale as the dynamical variables at the level above.

minimization. In formal terms, they map to its two components: the minimization of divergence and the maximization of evidence, see Fig. 2. Recognizing that action and perception can be unified within a single formal imperative – the minimization of free energy – is one of the key innovations of active inference, which helps integrate and extend the early theories reviewed above.

Regarding neural implementation, one of the most widely entertained hypotheses – about how the brain might implement perceptual inference – is predictive coding (Rao & Ballard, 1999). Fig. 3 shows the architecture of a predictive coding scheme as it might manifest in the cerebral cortex. In this predictive coding network, inference is realized by propagating predictions and prediction errors through top-down and bottom-up pathways, respectively, and by minimizing prediction errors across all levels. Interesting, predictive coding can be derived as a special case of variational free energy minimization (Friston, 2005).

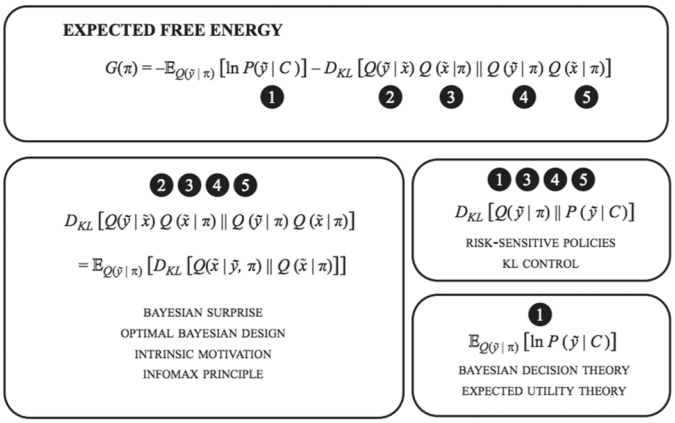

Fig. 4. Expected free energy and the way it can be mapped to different formal notions (e.g., Bayesian surprise, Risk-sensitive control, Expected utility theory) by removing one or more terms, denoted with numbers. This Figure—reproduced from (Parr et al., 2022)—expresses expected free energy in terms of beliefs about trajectories (indicated by the tilde ~). The additional symbols here, not in previous figures, are the π for policies and the *C* for preferences. Note that some terms (including term 1) are expressed in terms of expectations—i.e., averages under the subscripted distribution.

While predictive coding is a model of perception, it can be readily extended to encompass the role of action in the minimization of free energy (described above). The move from predictive coding to active inference can be realized by equipping predictive coding networks with simple motor reflexes. In this perspective, the motor system works by generating proprioceptive predictions (in the same way standard predictive coding generates exteroceptive predictions) — and not motor commands, as conventionally proposed – and these proprioceptive predictions are realized through the motor reflexes (Adams et al., 2013).

Subsequently, this theory was extended to also model autonomic control (Barrett & Simmons, 2015; Pezzulo, 2014; Seth et al., 2012). The general idea is that autonomic control might work by generating interoceptive predictions (i.e., homeostatic setpoints) and then fulfilling them through autonomic reflexes, in much the same way motor control might work by generating proprioceptive predictions and then fulfilling them through motor reflexes. This development of active inference helps connect it with theories of allostatic control (Sterling, 2012) and paves the way to a better understanding of our ability to model and control the internal milieu, not just the external environment. This stream of research underwrote novel approaches to psychopathology – as deficits of interoceptive processing (Paulus et al., 2019).

So far, we have discussed active inference using generative models that characterize processes that unfold in continuous time (e.g., predictive coding networks) and use continuous variables (i.e., the formal framework of dynamical systems and state-space models). However, many cognitive problems can be characterized at a distinct level: as (sequences of) discrete decisions. These include problems that require the selection of discrete responses during psychology experiments, the targets for saccades, or navigational trajectories in discretized environments (Friston et al., 2017; Friston, Lin, et al., 2017). These problems can be modeled in active inference, using generative models that use discrete variables (and the formal framework of Partially Observable Markov Decision Processes).

In addition to the two aforementioned components (priors and likelihood function), the generative models for active inference in discrete time often include a third component: the *transition function*, which describes the way in which hidden states change depending upon the agent's actions (or sequences of actions, called policies). Crucially, these generative models have temporal depth and afford a novel capability that was not available in simpler models: namely, planning. In simple terms, planning involves using the generative model to predict the consequences of different policies, scoring the policies according to how much they are *expected* to minimize free energy in the future and then (with some simplifications) select the best policy.

This planning process induces a novel quantity – *expected free energy* – that is the functional that active inference uses to evaluate (and assign a prior to) policies and it is distinct from the notion of variational free energy discussed so far (Friston et al., 2017). The notion of expected free energy has been very useful in the development of active inference models of things like (bounded) decision-making, planning, exploration-exploitation and curiosity (Friston, Lin, et al., 2017; Parr & Pezzulo, 2021; Schwartenbeck et al., 2019). This is because this notion is richer than the common optimization objectives used in other formal frameworks (e.g., economic theory and reinforcement learning). This is because expected free energy considers jointly a *pragmatic imperative* (utility maximization) and an *epistemic imperative* (information gain, or the resolution of the uncertainty). Indeed, as Fig. 4 illustrates, it is possible to map expected free energy to various other formal notions (e.g., Bayesian surprise, Risk-sensitive control, Expected utility theory), by removing one or more of its terms.

Active inference is a general scheme that can be applied to address various cognitive processes. Crucially, the functioning of active inference is the same across all problems: what differs is the generative model, which is task specific. This implies that by designing the appropriate generative models, it is possible to address a variety of cognitive tasks with the same approach – and to pass from the normative perspective of active inference to specific implementations that have biological plausibility (Friston, Parr, et al., 2017; Parr & Friston, 2018).

Here, a worked example may be helpful. To illustrate some of the principles we have outlined so far, we will consider how we might go about developing a model for a ubiquitous task in cognitive neuroscience—a delay period oculomotor task. This is a relatively simple task that can be performed by humans—and some animals—and that is designed to probe working memory function (Funahashi et al., 1989). The task sequence is as follows. First, a cross is presented on screen and our subject maintains fixation on this cross. A target then appears at one of several possible locations towards the periphery of the screen, but our subject still maintains fixation. The target then disappears and, after a 'delay period', a stimulus appears to signify that the subject should make a saccadic eye movement to the location of the target. Successful performance of this task relies upon retaining a memory of the target location during the delay and response phases.

To model this task, we must consider the data available to the subject. In this case, these are the visual stimuli and proprioceptive inputs, and whether the correct action was chosen. To do so, we need to take account of the causes of these data. The causes of proprioceptive data are simply the direction in which our subject's eyes are pointing. Visual outcomes, depend upon a combination of (1) gaze direction, (2) the intended target location, and (3) the current stage of the task (i.e., the fixation, target presentation, delay, or response stage). For each of these three variables, we must then specify how we expect them to evolve throughout the task. The gaze direction will transition from one step to the next based upon the decisions our subject makes. The intended target location will be fixed (although initially unknown) throughout the task. The task stage evolves predictably through a sequence of steps. Together, these beliefs about the way in which data are generated and the dynamics of the causes allow our subject to predict what will be observed next, and to update these beliefs when these predictions are violated.

As outlined above, active inference equips models with prior beliefs about the relative plausibility of different choices based upon their relative expected free energies. In this model, the key part of the expected free energy is a preference for receiving the 'correct' feedback outcome which is only available during the response phase of the task (see (Mirza et al., 2016) for a similar setup in the context of scene categorisation, in which the main role of the expected free energy is to promote information seeking). It is this that prompts a saccade to the

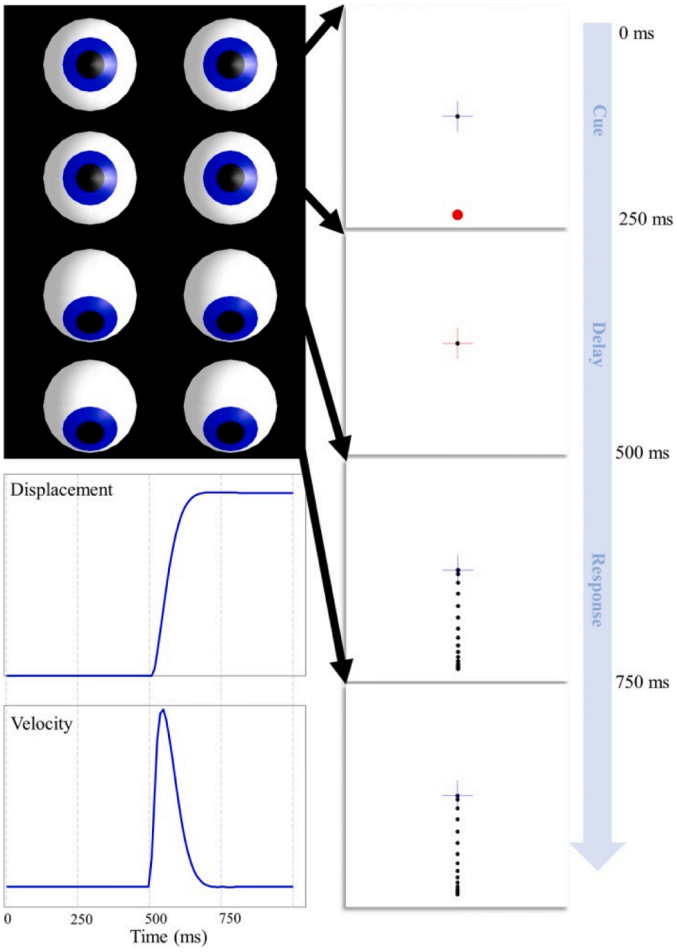

Fig. 5. A simulated oculomotor delay period task. This figure, taken from (Parr & Friston, 2019b) (published under a CC BY 4.0 license), shows simulated performance of a simple working memory task under active inference. Although simple, this task calls for planning (of our next saccade), recall (of the target location), and movement execution. The upper left images show a series of frames taken from the simulation, as if we were observing our participant's eyes. The black arrows link these behavioural responses to the view of the stimulus screen from the time of the target (red) presentation to the response phase. A series of black dots show the (cumulative) trajectory of gaze direction. Because this model is formulated to have both continuous (prediction-error minimising) and discrete (sequential planning) parts, we can plot the trajectory both in terms of position and velocity (lower left) and in terms of the sequence of actions taken.

remembered target location. Finally, the predicted action must be executed. This depends upon resolving the error between the anticipated proprioceptive information given the inferred saccade and current proprioceptive input. The result is the sequence of steps shown in Fig. 5.

The oculomotor control example illustrates how active inference can be concretely applied to study cognitive tasks, by designing (or learning) the appropriate generative models. Generative models represent formal hypotheses about how cognitive tasks are accomplished – hypotheses that can be validated with empirical data. A useful illustration of the design principles to realize (or train) generative models for different cognitive problems is provided in (Parr et al., 2022). This treatment makes a distinction between generative models in continuous time (that are useful to address motor control tasks) and discrete time (that are useful to address decision and planning tasks) and explains how these two types can be combined to form so-called hybrid or mixed generative models, in which discrete-time models are placed on top of continuous-time models. Furthermore, the generative models of active inference can be extended hierarchically, to model processes that unfold at different timescales. One example is the model of active listening processes, in which (for example) lower hierarchical levels deal with words and higher levels deal with sentences (Friston et al., 2021). Another example is a model of hierarchical action recognition that recognizes actions at different levels, from low level kinematics to higher level goals and intentions (Proietti et al., 2023). It is also possible to use hierarchical models to model hierarchies of control, in which lower-to-higher levels deal with autonomic imperatives (e.g., ensure a correct basic temperature) in increasingly complex ways (e.g., from thermoregulation to the goal-directed plan to buy water before a long run) (Pezzulo et al., 2015; Tschantz et al., 2021). These developments – from simple to more sophisticated (e.g., hierarchically and temporally deep) generative models has extended the range of cognitive models that have been addressed using active inference over the years.

Another interesting realization is the fact that it is possible to derive a biologically motivated "process theory" for active inference in discrete time, by interpreting the specific operations (variational updates) required to minimize free energy as signals that are computed or exchanged across neurons (Friston et al., 2017). This is important because it permits crossing levels of explanation – from normative to mechanistic and neuronal – and to use active inference to simulate neuronal activity that would ensue from the performance of cognitive tasks (Friston, Parr, et al., 2017; Parr & Friston, 2018).

Another important development of active inference regards precision control and its role in psychopathologies. In predictive coding, variables are encoded as Gaussian distributions and precision simply refers to the inverse of the variance of a distribution (Friston, 2005). Precision control refers to a mechanism that optimizes the precision of (the distribution of) each variable of an organism's generative model. It is important since it regulates the relative importance of top-down predictions and bottom-up prediction errors across the hierarchy. This is because prediction errors that are assigned greater (lower) precision have greater (lower) impact on the belief updating and the ensuing inference. Veridical inference requires the precision of (the distribution of) each variable to be optimized, to reflect the signal-to-noise ratio of sensory signals – therefore highlighting a link between precision control and attention as gain control (Parr & Friston, 2019a) – or the importance of an organism's prior preferences – reflecting the fact that an organism's innate drives or goal states can be encoded as highly precise priors (Pezzulo et al., 2015). Interestingly, when precision control fails, it can produce excessively rigid forms of inference (when priors fail to be updated in the light of novel evidence) or excessive sensitivity to stimuli (when belief revision follows the sensory input and its random fluctuations too closely—i.e., it overfits). These forms of aberrant inference, which depend sensitively on predicted precision, have been adopted to explain several psychopathological conditions, such as delusions, depression, psychosis, and many others (Barrett et al., 2016; Corlett & Fletcher, 2015; Edwards et al., 2012). In turn, these theories also speak to aberrant neuromodulation, since the precision of (the distribution of) different variables might be encoded by different neuromodulators, e.g., acetylcholine for the precision of the likelihood, noradrenaline for the precision of transitions, dopamine for the precision of policies, etc. (Parr & Friston, 2018).

Yet another development regards the analysis of generative models during sleep or other 'offline' periods. It has long been hypothesized that learning generative models benefits from alternating on-line and off-line periods (Hinton et al., 1995). While on-line generative modelling maximises accuracy (under complexity constraints), during off-line activity – in the absence of sensory data to "explain away" – model optimisation can focus on minimising complexity; for example, by removing redundant parameters (Friston, Lin, et al., 2017; Pezzulo et al., 2021). From a neuronal perspective, generative modelling during offline periods could be associated with (generative) replay activity in the hippocampus, the prefrontal cortex and other brain areas; but these links remain to be fully established (Foster, 2017; Schwartenbeck et al., 2023; Stoianov et al., 2022).

Finally, an interesting development regards the realization of active inference in which the free energy minimization extends "beyond the skull", to model the ways multiple active inference agents engage in cooperative or competitive tasks (Friston & Frith, 2015; Maisto et al., 2023) or construct their own niches (Constant et al., 2022). These and other works illustrate that the concept of free energy minimization can readily extend to multi-agent settings – including settings that go beyond the standard scope of cognitive science, such as morphogenesis (Friston & Levin, Sengupta, et al., 2015) and autopoiesis (Friston, 2013) – and hence potentially shed light on the relations between multiple nested levels of (self-)organization, from individual to social and cultural levels.

In sum, we have highlighted various developments of active inference, which encompass the complementary roles of perception and action in minimizing an organism's variational free energy (and ensuring that it successfully avoids "surprising" and characteristic states), the proposal of biologically plausible architectures for continuous time predictive coding and action control, the realization of generative models for discrete decisions that afford planning and the minimization of expected free energy, the hierarchical extension of these models, the importance of precision control, and beyond. For each of these topics, we have cited some selected papers that the interested readers might want to consult for more detailed information. Clearly, this is not an exhaustive list, but each of these developments has been useful to develop models of increasingly complex cognitive and social functions; see (Parr et al., 2022) for a more exhaustive treatment of active inference.

4. The benefits of unification

In the previous Section, we saw that the scope of active inference touches several domains of psychology and neuroscience. Here, we foreground a benefit of this rapid expansion: namely, unification.

Arguably, a main goal of cognitive psychology and neuroscience is explaining behavior and its neural foundations, in a comprehensive (if not a unified) way. Yet, to ensure methodological rigor, these disciplines usually adopt restricted laboratory settings that tend to isolate cognitive functions and obfuscate their relations (Maselli et al., 2023). Consider for example a mundane task that we solve almost every day: crossing a busy road. Even this relatively simple task engages several cognitive processes in a coordinated manner, such as perception (of the situation), memory (of past street crossing episodes), planning and action selection (of the best route), motivation (and the "why" of crossing), attention (to select the most relevant stimuli), etc. These processes are often studied in isolation using different paradigms leading to a proliferation of hypothesis and theories that assign each of them a distinct computational objective (and perhaps brain area) – therefore determining a very fragmented theoretical landscape.

Active inference proceeds the other way around: it starts from a single principle and asks how far one can go with it. And to what extent it is possible to derive from that principle empirically testable hypotheses about behavior and its cognitive and neural mechanisms? This approach brings the benefits of unification, in at least six ways.

First, active inference assumes that everything, from perception to action selection and learning ultimately serves to minimize variational free energy. A consequence of this is that one can align the (sometimes vague) conceptual terms used in psychology with crisp formal terms of free energy minimization. For example, one can assign things like attention to precision control. At the neuronal level, the fast updates – mediated by synaptic activity – might correspond to inferential processes that minimize free energy at a fast time scale, whereas the slower updates – at the level of synaptic efficacy – might correspond to learning processes that minimize free energy at a slower timescale. Precision dynamics might correspond to the activity of neuromodulators, which finesse the inference at multiple levels, for example, by increasing the post-synaptic gain of sensory or prediction error-units (Feldman & Friston, 2010). Oscillatory dynamics that are ubiquitous (and that often occur in synchrony) both within and across brain area might be signatures of temporal prediction and of the exchange of top-down and bottom-up information across hierarchical levels of the brain's generative model (Arnal & Giraud, 2012).

Second, active inference suggests that cognitive functions – usually addressed in isolation – might be instead better understood by appealing to a unique process theory. For example, in prominent computational neuroscience theories, perception and action are two separate functions with different objectives and neural substrates. According to Bayesian decision theory (Robert, 2007), the goal of perception is to provide an accurate estimate of the agent's state, whereas the goal of action selection is to maximize its expected utility. The former process is a precondition for the latter, implying an outdated, serial view of cognitive processing. Active inference holds that perception and action cooperate to minimize free energy, by minimizing divergence and maximizing evidence, respectively (Parr et al., 2022). Another example is the fact that in 20th-century cognitive science, working memory was considered as a separate storage that can be assessed by other components when needed; therefore, imposing a separation between information storage and information processing. In contrast, active inference models of hierarchical perception and action (Friston et al., 2021; Pezzulo et al., 2018) treat memory of the previous state as intrinsic to the belief updating under generative or world models, across multiple timescales, which is in keeping with 21st-century accounts of working memory (Hasson et al., 2015).

Third, active inference has the potential to unify different "levels of understanding" of cognitive processes. Marr famously introduced a distinction between computational, algorithmic and neural implementation levels and argued that progress can be made within each level and by connecting different levels (Marr, 1982). Establishing links between theories that operate at different levels is often challenging. Active inference helps establish firm relations across levels of description. Rather than Marr's tripartite distinction, in active inference it is more common to appeal to a distinction between *normative theory* and *process theory* (Friston et al., 2017). Free energy minimization is the normative objective of living organisms, whereas predictive coding and variational message passing are process-level theories that describe how the brain might support free energy minimization. Importantly, as shown by (Friston, 2005), under certain assumptions predictive coding can be directly derived by the minimization of variational free energy, connecting the two levels of explanation. A similar case can be made for the variational message passing schemes proposed to support discrete active inference in neural circuits (Friston et al., 2017).

Fourth, unification endows existing constructs with validity, via the application of active inference across domains. One example is the development of theories of interoceptive inference and autonomic control (Barrett & Simmons, 2015; Pezzulo, 2014; Seth et al., 2012) by analogy with the functioning of action control (Adams et al., 2013). In this perspective, autonomic control works exactly like action control – namely, it aims to minimize a discrepancy between a predicted and a sensed signal – except that the "signal" refers to interoceptive streams rather than proprioceptive streams. Another example can be found in computational psychiatry, where numerous accounts of psychopathology appeal to a single mechanism: namely, aberrant precision control.

Fifth, active inference has the potential to reconcile (or at least to contextualize) theoretical perspectives that have long been considered at odds in psychology, neuroscience and philosophy. One example is the Helmholtzian view that perception constitutes an inference about the entities of the external world that cause our sensations (Helmholtz, 1866) and the Gibsonian view that perceiving consists in seeing action opportunities and affordances, not reconstructing a model of the external reality within the brain (Gibson, 1979). This apparent dialectic could be dissolved by considering that there are multiple ways to design generative models; specifically, a relevant distinction is between generative models that explicitly model the ways external states produce

Box 1
Glossary of technical terms.

Active Inference: A normative framework that elucidates the neural and cognitive processes underlying sentient behavior, beginning with first principles. This framework posits that perception and action work in concert to minimize a shared functional known as variational free energy.

Expected Free Energy: This is the quantity that is used in active inference to score action sequences or policies (and then to select between them). It takes into consideration both the pragmatic value of policies – or how close a policy's expected outcomes are to the preferred outcomes – and their epistemic value (or information gain) – or how much the policy is expected to reduce uncertainty.

Generative Model: A statistical model designed to explain the generation of observable content from unobservable, hidden (latent) causes. For instance, it clarifies the process by which a visual object gives rise to an image on the retina. Generative models serve a dual purpose: they allow the generation of novel, synthetic content and support the inference of hidden causes from observable data. From a technical standpoint, generative models encode the joint probability distribution governing both observables and hidden causes.

Latent (or Hidden) Variable: An internal variable within a generative model, referred to as "latent" or "hidden" due to the fact that it cannot be directly observed, but must be inferred.

Precision and precision-weighting: Precision denotes the inverse of variance or standard error, serving as a measure of the reliability or certainty associated with sensory information. Precision-weighting refers to the fact that in predictive coding and active inference, prediction errors are weighted by their respective precisions, therefore determining the extent to which sensory observations influence the process of updating beliefs.

Predictive Coding: A computational framework in neuroscience that provides a possible neural implementation for the idea that perception consists in a process of inference. In hierarchical predictive coding networks, inference is realized by minimizing (precision-weighted) prediction errors across all hierarchical levels. In turn, this requires bidirectional loops between top-down processes (conveying predictions) and bottom-up processes (conveying prediction errors).

Variational Free Energy: This is the functional (function of a function) that is minimized within the framework of active inference. It is also widely utilized in utilized in probabilistic modeling, statistical inference and machine learning. In its simplest instantiation, it corresponds to a summation of prediction errors, which quantifies the deviation of observed data from the predictions of the generative model. More formally, variational free energy serves as an upper bound on the negative logarithm of the evidence, which is the probability of observed data given a model.

sensations (a.k.a., environmental models) or the ways actions produce sensations (a.k.a., sensorimotor models) (Sims & Pezzulo, 2021; Pezzulo et al., 2023). Some active inference studies use generative models that include explicit beliefs about entities in the external world that cause sensations, such as one's location in space (Friston et al., 2017). Other active inference studies use generative models that only consider the sensory consequences of one's action, such as touch sensations that follow whisking at a given amplitude, but not explicit beliefs about objects 'out there' (Mannella et al., 2021). The latter generative models adhere more closely to the notions of affordance (Gibson, 1979) and of sensorimotor contingency (O'Regan & Noe, 2001), despite the fact they still entail inferential dynamics. Besides this specific topic, there is a vivid debate in philosophy that concerns the most appropriate way to consider active inference, in relation to internalist (Hohwy, 2013), externalist (Clark, 2013) or enactivist theories (Bruineberg et al., 2018).

Finally, and importantly, the integrative perspective of active inference could be valuable in characterising of sentient behaviour – considered here to be the capacity to infer states of the world and to act upon it with a sense of purpose (Friston, Da Costa, et al., 2023). This operational definition is satisfied by active inference when, and only when the generative model includes the consequences of action (mathematically, when the generative model includes priors over policies based upon expected free energy). This notion of sentience is does not have any phenomenological commitments and is probably best read as 'basic sentience' in the sense of (Clark, 2023).

Recently, there has been a proliferation of advanced Generative AI systems that process language, images and videos with very high accuracy. However, in most cases, these systems learn passively from large predefined datasets and disregard agency – and the possibility to act upon the world with a purpose – to develop genuine understanding (Pezzulo et al., 2023). Active inference suggests a different path to understand and simulate sentient behaviour, which focuses on the development of grounded world (i.e., generative) models, by actively engaging with the environment and by predicting the consequences of the requisite interactions. An open question for future research is whether the enactive and embodied approach of active inference has the potential to complement and advance the development and deployment of Generative AI.

5. Opportunities for the future

It's Difficult To Make Predictions, Especially About the Future. Niels Bohr.

The compass of active inference is expanding rapidly, but the landscape of future opportunities may be even ampler. Here, we focus on some of the developments that we consider most promising and most likely in the near future.

The first and perhaps most obvious direction for the future regards a deeper empirical scrutiny of active inference. A question that is sometimes asked of active inference is whether any empirical findings could offer evidence for or against the framework. This can be a vexed question to answer as it constitutes a category error. A framework is not in itself a hypothesis. It is a way of formulating hypotheses. The relationship between active inference and empirical psychology is that we can formalize psychological theories in terms of the generative models that underwrite neurophysiological and behavioural responses. Equipped with a proposed model, the framework can be used to express a hypothesis, to predict the behaviour expected under that hypothesis, and to fit to measured data to formally compare alternative hypotheses. In other words, while active inference is an application of the free energy principle – which is a principle (i.e., method) rather than a theory (Friston, 2010) – theories tested under the active inference framework (e.g., those considered in this article) make specific empirical predictions that can (and need to) be empirically validated. One example of this is the oculomotor delay period model shown in Fig. 5, which generate empirically testable predictions about oculomotor performance as a function of varying delay periods (Parr & Friston, 2019b). Various empirical studies are already addressing the empirical

predictions of predictive coding, such as how top-down and bottom-up dynamics support predictions and prediction errors, respectively (Walsh et al., 2020). However, active inference makes a number of specific predictions about (for example) the way the motor system works (Shipp et al., 2013) and the way higher cognitive functions are implemented (Pezzulo et al., 2018) that differ from mainstream theories and could be increasingly scrutinized by future studies.

A second interesting direction for the future is assessing to what extent active inference – and more broadly, the free energy principle – can help us understand the evolution of complex neural circuits and life forms from simpler ones. Active inference suggests a possible path from the simple mechanisms that supported prediction and control in our earlier evolutionary ancestors to the more sophisticated abilities of our species (Pezzulo et al., 2022), but a comprehensive account of the evolution and "phylogenetic refinement" (Cisek, 2019) of living organisms remains to be fully developed (Friston et al., 2023; Friston, Friedman, et al., 2023).

A third interesting direction for the future regards the realization of advanced artefacts, such as AIs and robots, based on active inference. There have already been several successful robotic implementations of active inference, but the full potential of the theory has not yet been reached (Ahmadi & Tani, 2019; Lanillos et al., 2021; Priorelli et al., 2023; Taniguchi et al., 2023). Interestingly, some of the central concepts of active inference, such as the importance of generative models and self-supervised, predictive learning, are becoming central in mainstream research in AI, as testified by the recent successes in generative AIs such as large language models. This creates an important opportunity, since (apart for their obvious technological impact), state-of-the-art AI systems can be precious in advancing our understanding of living organisms, providing that they incorporate appropriate (design) principles (Pezzulo et al., 2023).

Funding and acknowledgements

This research received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreements No. 945539 (Human Brain Project SGA3) to GP and KF and No. 952215 (TAILOR) to GP; the European Research Council under the Grant Agreement No. 820213 (ThinkAhead) to GP; the PNRR MUR projects PE0000013-FAIR and IR0000011–EBRAINS-Italy to GP; for the Wellcome Centre for Human Neuroimaging (Ref: 205103/Z/16/Z) to KF, a Canada-UK Artificial Intelligence Initiative (Ref: ES/T01279X/1) to KF. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors did not use generative AI technologies for preparation of this work.

Declaration of Competing Interest

The authors declare no conflict of interest.

Data availability

No data was used for the research described in the article.

References

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure & Function, 218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Ahmadi, A., & Tani, J. (2019). A novel predictive-coding-inspired variational RNN model for online prediction and recognition. *Neural Computation, 31*(11), 2025–2074. https://doi.org/10.1162/neco_a_01228

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16*(7), 390–398. https://doi.org/10.1016/j.tics.2012.05.003

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience, 16*(7), 419–429. https://doi.org/10.1038/nrn3950

Barrett, L. F., Quigley, K. S., & Hamilton, P. (2016). An active inference theory of allostasis and interoception in depression. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1708), 20160011. https://doi.org/10.1098/rstb.2016.0011

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695–711.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences, 16*(10), 485–488. https://doi.org/10.1016/j.tics.2012.08.006

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2018). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese, 195*(6), 2417–2444. https://doi.org/10.1007/s11229-016-1239-1

Buzsaki, G. (2019). USA. *The brain from inside out*. Oxford University Press.

Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological Reviews, 9*(3), 399–431.

Cisek, P. (1999). Beyond the computer metaphor: behaviour as interaction. *Journal of Consciousness Studies, 6*(11–12), 11–12.

Cisek, P. (2019). Resynthesizing behavior through phylogenetic refinement. *Attention, Perception, & Psychophysics, 81*(7), 2265–2287. https://doi.org/10.3758/s13414-019-01760-1

Cisek, P., & Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neu, 33*, 269–298. https://doi.org/10.1146/annurev.neuro.051508.135409

Cisek, P., & Pastor-Bernier, A. (2014). On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society B*.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(03), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press, Incorporated.

Clark, A. (2023). *The experience machine: How our minds predict and shape reality*.

Conant, R. C., & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 89–97.

Constant, A., Clark, A., Kirchhoff, M., & Friston, K. J. (2022). Extended active inference: Constructing predictive cognition beyond skulls. *Mind & Language, 37*(3), 373–394. https://doi.org/10.1111/mila.12330

Corlett, P. R., & Fletcher, P. C. (2015). Delusions and prediction error: Clarifying the roles of behavioural and brain responses. *Cognitive Neuropsychiatry, 20*(2), 95–105. https://doi.org/10.1080/13546805.2014.990625

Craik, K. (1943). *The Nature of Explanation*. Cambridge University Press.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (Eds.). (2007). *Bayesian Brain: Probabilistic Approaches to Neural Coding* (1st ed..,). The MIT Press. http://www.amazon.com/exe%bidos/redirect?tag=citeulike07-20&path=ASIN/026204238X.

Edwards, M. J., Adams, R. A., Brown, H., Pareés, I., & Friston, K. J. (2012). A Bayesian account of "hysteria. *Brain: A Journal of Neurology, 135*(Pt 11), 3495–3512. https://doi.org/10.1093/brain/aws129

Engel, A. K., Friston, K. J., & Kragic, D. (2016). *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science*. MIT Press.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience, 4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Foster, D. J. (2017). Replay comes of age. *Annual Review of Neuroscience, 40*, 581–602.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K. (2013). Life as we know it. *Journal of The Royal Society Interface, 10*(86). https://doi.org/10.1098/rsif.2013.0475

Friston, K., & Frith, C. (2015). A Duet for one. *Consciousness and Cognition, 36*, 390–405.

Friston, K., Parr, T., & de Vries, B. (2017). The graphical brain: Belief propagation and active inference. *Network Neuroscience (Cambridge, Mass), 1*(4), 381–414. https://doi.org/10.1162/NETN_a_00018

Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biol Cybern, 102*(3), 227–260. https://doi.org/10.1007/s00422-010-0364-z

Friston, K., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place: A free-energy approach to pattern regulation. *Journal of The Royal Society Interface, 12*(105), 20141383. https://doi.org/10.1098/rsif.2014.1383

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Computation, 29*(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Friston, K., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017). Active Inference, Curiosity and Insight. *Neural Computation, 29*(10), 2633–2683. https://doi.org/10.1162/neco_a_00999

Friston, K., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., & Holmes, E. (2021). Active listening. *Hearing Research, 399*, Article 107998. https://doi.org/10.1016/j.heares.2020.107998

Friston, K., Friedman, D.A., Constant, A., Knight, V.B., Parr, T., & Campbell, J.O. (2023). *A variational synthesis of evolutionary and developmental dynamics*. https://doi.org/10.3390/e25070964.

Friston, K., Da Costa, L., Sakthivadivel, D. A. R., Heins, C., Pavliotis, G. A., Ramstead, M., & Parr, T. (2023). Path integrals, particular kinds, and strange things (arXiv: 2210.12761). *arXiv*. https://doi.org/10.48550/arXiv.2210.12761

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61*(2), 331–349. https://doi.org/10.1152/jn.1989.61.2.331

Gibson, J. J. (1979). *The ecological approach to visual perception*. Lawrence Erlbaum Associates, Inc.
Gregory, R. L. (1968). Perceptual illusions and brain models. *Proceedings of the Royal Society of London Series B Biological Sciences, 171*(1024), 279–296.
Gregory, R. L. (1980). Perceptions as Hypotheses. *Philosophical Transactions of the Royal Society of London B, Biological Sciences, 290*(1038), 181–197. https://doi.org/10.1098/rstb.1980.0090
Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences, 19*(6), 304–313. https://doi.org/10.1016/j.tics.2015.04.006
Helmholtz, H. von (1866). Concerning the perceptions in general. In J. P. C. Southall (Ed.), *Treatise on physiological optics* (Vol. 3). Dover.
Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The "wake-sleep" algorithm for unsupervised neural networks. *Science, 268*(5214), 1158–1161.
Hoffmann, J. (2003). Anticipatory behavioral control. In M. V. Butz, O. Sigaud, & P. Gerard (Eds.), *Anticipatory Behavior in Adaptive Learning Systems: Foundations, Theories, and Systems* (pp. 44–65). Springer-Verlag.
Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
Hommel, B. (2003). Planning and representing intentional action. *TheScientificWorld JOURNAL, 3*, 593–608.
James, W. (1890). *The Principles of Psychology*. Dover Publications.
Lanillos, P., Meo, C., Pezzato, C., Meera, A. A., Baioumy, M., Ohata, W., Tschantz, A., Millidge, B., Wisse, M., Buckley, C. L., & Tani, J. (2021). *Active Inference in Robotics and Artificial Agents: Survey and Challenges* (arXiv:2112.01871). *arXiv*. https://doi.org/10.48550/arXiv.2112.01871
Lepora, N. F., & Pezzulo, G. (2015). Embodied Choice: How Action Influences Perceptual Decision Making. *PLoS Comput Biol, 11*(4), Article e1004110. https://doi.org/10.1371/journal.pcbi.1004110
Maisto, D., Donnarumma, F., & Pezzulo, G. (2023). Interactive Inference: A Multi-Agent Model of Cooperative Joint Actions. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 1–12. https://doi.org/10.1109/TSMC.2023.3312585
Maisto, D., Barca, L., Van den Bergh, O., & Pezzulo, G. (2021). Perception and misperception of bodily symptoms from an active inference perspective: Modelling the case of panic disorder. *Psychological Review, 128*(4), 690–710. https://doi.org/10.1037/rev0000290
Mannella, F., Maggiore, F., Baltieri, M., & Pezzulo, G. (2021). Active inference through whiskers. *Neural Networks: The Official Journal of the International Neural Network Society, 144*, 428–437. https://doi.org/10.1016/j.neunet.2021.08.037
Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt and Co., Inc.. http://portal.acm.org/citation.cfm?id=1096911.
Maselli, A., Gordon, J.R., Eluchans, M., Lancia, G.L., Thiery, T., Moretti, R., Cisek, P., & Pezzulo, G. (2023). *Beyond simple laboratory studies: Developing sophisticated models to study rich behavior*.
Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of Living*. D. Reidel Pub.
Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the Structure of Behavior*. Holt, Rinehart and Winston.
Mirza, M. B., Adams, R. A., Mathys, C. D., & Friston, K. J. (2016). Scene Construction, Visual Foraging, and Active Inference. *Frontiers in Computational Neuroscience, 10*, 56. https://doi.org/10.3389/fncom.2016.00056
O'Regan, J. K., & Noe, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*(5), 883–917.
Parr, T., & Friston, K. J. (2018). The Anatomy of Inference: Generative Models and Brain Structure. *Frontiers in Computational Neuroscience, 12*. https://www.frontiersin.org/articles/10.3389/fncom.2018.00090.
Parr, T., & Friston, K. J. (2019a). Attention or salience? *Current Opinion in Psychology, 29*, 1–5. https://doi.org/10.1016/j.copsyc.2018.10.006
Parr, T., & Friston, K. J. (2019b). The computational pharmacology of oculomotion. *Psychopharmacology, 236*(8), 2473–2484. https://doi.org/10.1007/s00213-019-05240-0
Parr, T., & Pezzulo, G. (2021). Understanding, Explanation, and Active Inference. *Frontiers in Systems Neuroscience, 15*. https://www.frontiersin.org/articles/10.3389/fnsys.2021.772641.
Parr, T., Pezzulo, G., & Friston, K. J. (2022). Active Inference: The Free Energy Principle in. *Mind, Brain, and Behavior*. MIT Press.
Parr, T., Limanowski, J., Rawji, V., & Friston, K. (2021). The computational neurology of movement under active inference. *Brain, 144*(6), 1799–1818. https://doi.org/10.1093/brain/awab085
Parr, T., Holmes, E., Friston, K. J., & Pezzulo, G. (2023). Cognitive effort and active inference. *Neuropsychologia, 184*, Article 108562. https://doi.org/10.1016/j.neuropsychologia.2023.108562
Paulus, M. P., Feinstein, J. S., & Khalsa, S. S. (2019). An Active Inference Approach to Interoceptive Psychopathology. *Annual Review of Clinical Psychology, 15*(1), 97–122. https://doi.org/10.1146/annurev-clinpsy-050718-095617
Pezzulo, G. (2014). Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cognitive, Affective & Behavioral Neuroscience, 14*(3), 902–911. https://doi.org/10.3758/s13415-013-0227-x
Pezzulo, G., & Cisek, P. (2016). Navigating the Affordance Landscape: Feedback Control as a Process Model of Behavior and Cognition. *Trends in Cognitive Sciences, 20*(6), 414–424. https://doi.org/10.1016/j.tics.2016.03.013
Pezzulo, G., Rigoli, F., & Friston, K. J. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology, 136*, 17–35.
Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical Active Inference: A Theory of Motivated Control. *Trends in Cognitive Sciences, 0*(0). https://doi.org/10.1016/j.tics.2018.01.009
Pezzulo, G., Zorzi, M., & Corbetta, M. (2021). The secret life of predictive brains: What's spontaneous activity for? *Trends in Cognitive Sciences, 25*(9), 730–743. https://doi.org/10.1016/j.tics.2021.05.007
Pezzulo, G., Parr, T., & Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 377*(1844), Article 20200531. https://doi.org/10.1098/rstb.2020.0531
Pezzulo, G., Parr, T., Cisek, P., Clark, A., & Friston, K. (2023). *Generating Meaning: Active Inference and the Scope and Limits of Passive AI*.
Pezzulo, G., D'Amato, L., Mannella, F., Priorelli, M., Van de Maele, T., Stoianov, I. P., & Friston, K. (2023). Neural representation in active inference: using generative models to interact with–and understand–the lived world. arXiv preprint arXiv: 2310.14810.
Powers, W. T. (1973). *Behavior: The Control of Perception*. Aldine.
Priorelli, M., Pezzulo, G., & Stoianov, I. P. (2023). Deep kinematic inference affords efficient and scalable control of bodily movements. *Proceedings of the National Academy of Sciences, 120*(51), Article e2309058120.
Proietti, R., Pezzulo, G., & Tessari, A. (2023). An active inference model of hierarchical action understanding, learning and imitation. *Physics of Life Reviews, 46*, 92–118. https://doi.org/10.1016/j.plrev.2023.05.012
Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci, 2*(1), 79–87. https://doi.org/10.1038/4580
Rens, N., Lancia, G. L., Eluchans, M., Schwartenbeck, P., Cunnington, R., & Pezzulo, G. (2023). Evidence for entropy maximisation in human free choice behaviour. *Cognition, 232*, Article 105328. https://doi.org/10.1016/j.cognition.2022.105328
Robert, C. P. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation* (Vol. 2). Springer.
Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife, 8*, Article e41703. https://doi.org/10.7554/eLife.41703
Schwartenbeck, P., Baram, A., Liu, Y., Mark, S., Muller, T., Dolan, R., Botvinick, M., Kurth-Nelson, Z., & Behrens, T. (2023). Generative replay underlies compositional inference in the hippocampal-prefrontal circuit. *Cell*. https://doi.org/10.1016/j.cell.2023.09.004
Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00395
Shipp, S., Adams, R. A., & Friston, K. J. (2013). Reflections on agranular architecture: Predictive coding in the motor cortex. *Trends in Neurosciences, 36*(12), 706–716. https://doi.org/10.1016/j.tins.2013.09.004
Sims, M., & Pezzulo, G. (2021). Modelling ourselves: What the free energy principle reveals about our implicit notions of representation. *Synthese*. https://doi.org/10.1007/s11229-021-03140-5
Sterling, P. (2012). Allostasis: A model of predictive regulation. *Physiology & Behavior, 106*(1), 5–15.
Stoianov, I., Maisto, D., & Pezzulo, G. (2022). The hippocampal formation as a hierarchical generative model supporting generative replay and continual learning. *Progress in Neurobiology, 217*, Article 102329. https://doi.org/10.1016/j.pneurobio.2022.102329
Taniguchi, T., Murata, S., Suzuki, M., Ognibene, D., Lanillos, P., Ugur, E., Jamone, L., Nakamura, T., Ciria, A., Lara, B., & Pezzulo, G. (2023). World Models and Predictive Coding for Cognitive and Developmental Robotics: Frontiers and Challenges (arXiv: 2301.05832). *arXiv*. https://doi.org/10.48550/arXiv.2301.05832
Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*, 189–208.
Tschantz, A., Barca, L., Maisto, D., Buckley, C.L., Seth, A.K., & Pezzulo, G. (2021). Simulating homeostatic, allostatic and goal-directed forms of interoceptive control using Active Inference. *bioRxiv*, 2021.02.16.431365. https://doi.org/10.1101/2021.02.16.431365.
Van den Bergh, O., Witthöft, M., Petersen, S., & Brown, R. J. (2017). Symptoms and the body: Taking the inferential leap. *Neuroscience and Biobehavioral Reviews, 74*(Pt A), 185–203. https://doi.org/10.1016/j.neubiorev.2017.01.015
Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences, 1464*(1), 242–268. https://doi.org/10.1111/nyas.14321
Wiener, N. (1948). *Cybernetics: Or Control and Communication in the Animal and the Machine*. The MIT Press.