The Trident Brain Training System

Background Tutorials

Mark Ashton Smith, Ph.D.

University of Essex Online
Mindware Lab

February, 2024

I. The Bayesian Brain

The Bayesian Brain hypothesis is based on the idea that the brain is a statistical ‘inference machine’ that actively generates and updates hypotheses about the causes of its sensory input - the states of the world. These hypotheses take the form of internal models or representations of the world, which the brain uses to predict incoming sensory data. In this framework, perception is viewed as a process of hypothesis testing, where the brain uses its current model of the world to predict sensory input, then updates this model based on the discrepancy (*prediction error*) between the predicted and actual sensory data according to Bayes theorem below. Bayes' theorem is a fundamental principle in the field of probability theory and statistics that describes how to update the probabilities of hypotheses (in this case, beliefs or models of the world) when given evidence.

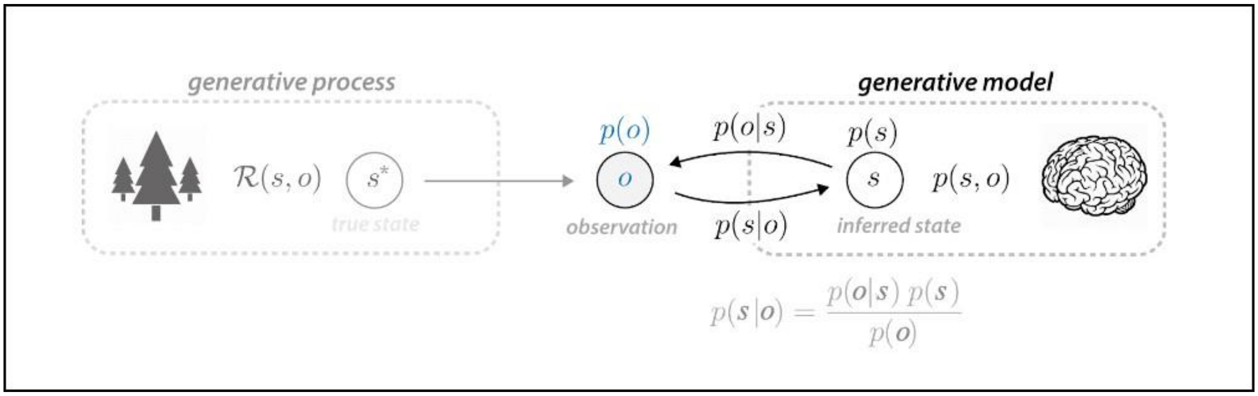
The theorem can be expressed as follows:

$$P(Hypothesis \mid Evidence) = \underbrace{P(Hypothesis)}_{\text{Prior}} * \frac{\overbrace{P(Evidence \mid Hypothesis)}^{\text{Likelihood}}}{\underbrace{P(Evidence)}_{\text{Normalizer}}}$$

Where:

P(H|E) is the posterior probability, or the probability of hypothesis H being true given the evidence E.
P(E|H) is the likelihood, or the probability of the evidence given that the hypothesis is true.
P(H) is the prior probability, or the initial degree of belief in H.
P(E) is the probability of the evidence.

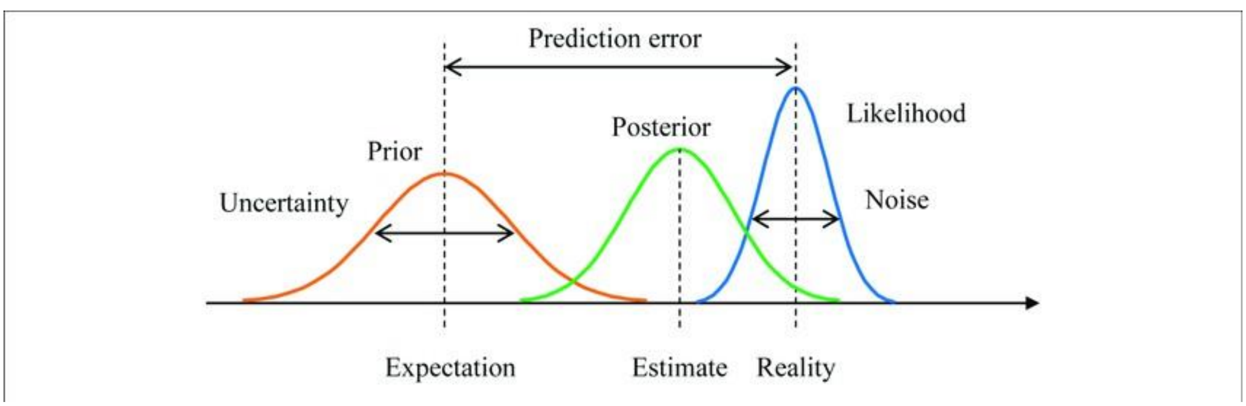
The Bayesian brain at work is shown in this figure below (where E = sensory observations ‘o’, and H = inferred states ‘s’).



Adapted from Oleg Solopchuk, 2018

Each term in Bayes' theorem (the prior, likelihood, and posterior) can be understood as a *probability distribution*, not just a discrete probability. Each distribution has its own mean and variance, which represent the average value and variance/uncertainty (or inverse variance/precision). So prediction error that drives the updating of internal beliefs is a function of both the difference between the prior model prediction and the sensory data (given the prediction), as well as the confidence level (precision) of the prediction and the signal to noise ratio (precision) of the sensory evidence.

These ideas are shown in the figure.



II. The Free Energy Principle (FEP)

The Free Energy Principle, first formulated by the British neuroscientist Karl Friston, extends the Bayesian Brain account in the following ways.

1. Free Energy (*F*): Prediction Error & Complexity-Simplicity Tradeoff

In traditional Bayesian accounts, the updating of beliefs (i.e., our internal models of the world) is based on the prediction error weighted by the level of confidence or certainty we have in our predictions and sensory data. The FEP maintains this principle, but claims that we strive to minimise *free energy* (*F*) - a quantity that is a function of both prediction error and the complexity of our internal generative model. ('Generative' in the sense of generating predictions about the sensory data we expect to encounter.)

Free energy is a measure of surprise or improbability of observed states in the world given our internal model of the causes of those states, and is minimised when the prediction error is low and the model is simple. The FEP Bayesian brain tries to achieve a balance between maintaining a model that is complex enough to accurately predict sensory data, while also adopting Occam's razor and being as simple as possible, which helps with efficiency. Simpler models conserve computational resources. They require less information to be stored and processed, allowing us to process information more efficiently. This is crucial given the brain's constraints in terms of energy usage and processing capacity.

2. Free Energy Formula

The formula for free energy (*F*) is:

$$F = -\log P(D|M) + D[Q(\theta|D, M) \,||\, P(\theta|D, M)]$$

The first term is the surprise/improbability of the sensory data (D) given the model (M). The more surprise, the more free energy. The second term is the 'KL divergence' which quantifies the dissimilarity or divergence between two probability distributions - $Q(\theta|D, M)$ and $P(\theta|D, M)$. (θ typically represents the model parameters.) The more the divergence, the more the free energy (*F*).

Q(θ|D, M) is the approximated posterior, representing our beliefs about the parameters θ of the model M given the data D. This is our updated belief about the state of the world (parameters) after receiving the sensory data. P(θ|D, M) is the true posterior, representing what the beliefs about the parameters θ should be if calculated exactly given the data D and the model M. This is generally intractable to compute exactly, which is why we have the approximation Q. The KL divergence term D[Q(θ|D, M) || P(θ|D, M)] thus measures how much our beliefs Q deviate from the true posterior P. This can be seen as a complexity cost: the more complex our model (i.e., the more parameters it has, the more detailed its representation of the world), the more potential there is for the approximate posterior Q to deviate from the true posterior P, resulting in a higher KL divergence.

So the Free Energy Principle captures Occam's razor by penalising models that are more complex than necessary to explain the data.

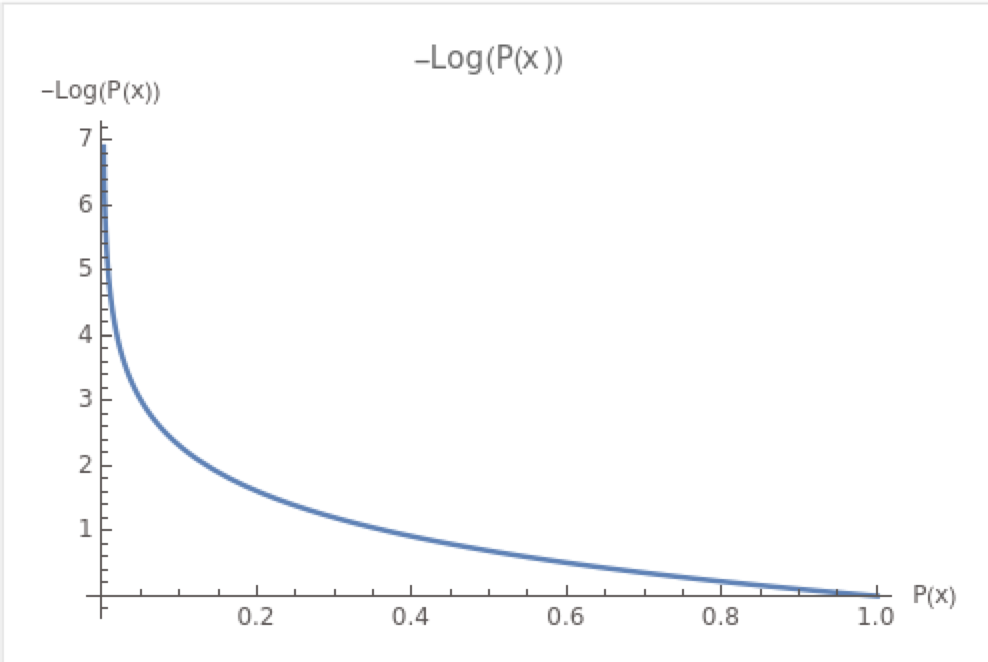
3. Free Energy (*F*) & Model Evidence

In Bayesian inference, when there is more evidence supporting a particular belief or hypothesis, the confidence - and thus precision - in that belief increases. As the weight of evidence increases, the model's predictions become more certain and less variable.
This extends to minimising free energy. According to the Free Energy Principle, an agent should select models and beliefs that minimise its free energy. As a model's evidence increases, the precision of its predictions increases, and thus its free energy decreases.

The relationship between free energy and model evidence is deeply embedded in the mathematics of the FEP and the Bayesian framework upon which it's built. The formula for free energy (*F*) is:

$$F = -\log P(D|M) + \text{KL divergence term}$$

Model evidence is denoted as P(D|M) where D is data and M is a model. This is the probability of observing the data given a particular model and quantifies how well the model predicts the data, giving a measure of the model's fit or explanatory power. High model evidence means that the model predicts the data well.



Plot of -Log(P(x))

As you can see, as the probability P(x) approaches 0, the value of −log(P(x)) increases sharply. Conversely, as the probability P(x) approaches 1, the value of −log(P(x)) approaches 0.

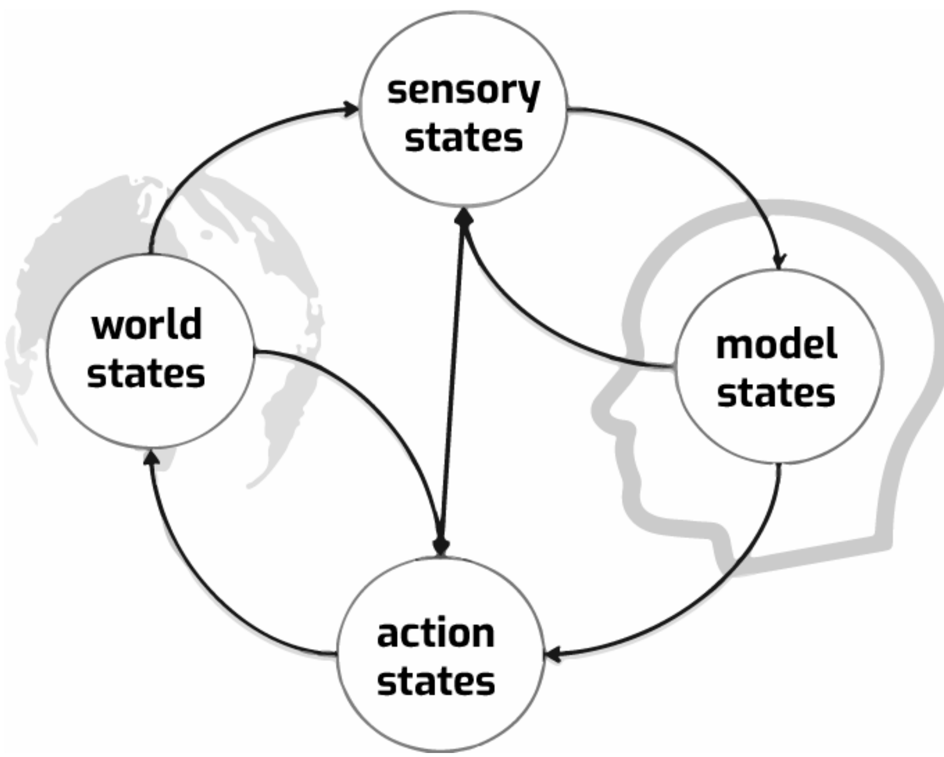
Free energy (*F*) is a functional that sets an upper limit on the negative log of the model evidence -log P(D|M). That is, the value of the free energy is always greater than or equal to the negative log of the model evidence - *F* ≥ -log P(D|M).

Functionals take functions as inputs (rather than numbers) and return numbers as outputs. In this case the function takes probability distributions of the model and data as inputs, and F can be understood as a 'summary statistic' of the approximated posterior P(DIM) or model evidence. *F* sets an upper limit on the negative log of this model evidence -log P(D|M). This simply means that the value of the free energy is always greater than or equal to the negative log of the model evidence, since the 'divergence' term' (see above) is always added to P(DIM).

Based on the formula, as evidence is increased, *F* is decreased. And if free energy is minimised where *F* = logP(DIM), this implies that model evidence is maximised, since we are minimising the negative log of the evidence. And generally, whenever …

4. Active Inference

The FEP extends the Bayesian Brain hypothesis by proposing that we not only update our beliefs in response to sensory input to reduce prediction error, but engage proactively with our environment to gain more information to reduce uncertainty and bring about states to conform with our predictions. The FEP puts *agency* and our ability to shape our worlds at centre stage.



Adapted from Sims & Pezzulo, 2021. The agent can change sensory states through action to reduce free energy in 'epistemic foraging' or expected free energy in enacting preferences and goal pursuit.

A. Expected Free Energy E(F): Pragmatic and Epistemic Active Inference

While actual free energy refers to a current measure of prediction error, uncertainty and model complexity, *expected free energy E(F)* concerns expectations about *future* free energy.

A ‘policy’ refers to planned actions - a sequence of actions that an agent can take to achieve a goal or realise a preference. E(F) is the future predicted surprise or uncertainty that would result from a specific policy, given our current model of the world. It incorporates both the expected **cost** (the difference between expected and preferred sensory outcomes) and the **uncertainty** about these outcomes.

Expected free energy is calculated for each potential policy that the agent could enact, and the one with the lowest expected free energy is selected. We attempt to minimise E(F) through a form of internal 'simulation', where we mentally play out the possible outcomes of different policies before selecting the one that is expected to minimise free energy (F) the most. This means that the agent is choosing policies and actions that it predicts will best reduce future uncertainty and surprise, effectively navigating the balance between the agent's goals and the need to reduce uncertainty about the world. The minimisation of expected free energy in this way leads to two types of active inference:

Pragmatic active inference. This is directed at fulfilling preferences or goals (i.e., reducing the expected cost/prediction error). For example, if the goal is to find our keys, we might select a policy that involves searching in areas where keys have been found before, based on our working model of the situation. This reflects an **exploitation** focused strategy - exploiting the current understanding we already have.

Epistemic active inference. This is aimed at reducing uncertainty and ambiguity. This might involve a policy of exploration to gather more information, even if it doesn't directly lead to the satisfaction of a specific goal or preference. For instance, we might adopt a policy of exploring an unfamiliar environment in order to build a more accurate and reliable model of that environment as we learn how to navigate it. Or we may collect more information on a topic before making a decision about it. This is called ‘epistemic foraging.’ It is an *exploration* focused strategy.

B. Goal State Directedness

In the formalism of the FEP, there is an inherent distinction between state and inference. States are configurations of the system at a specific point in time - for example, posterior beliefs about the state of the world. Inference, on the other hand, is the computational process through which the system updates its beliefs (i.e., changes its state).

In terms of free energy (*F*) minimisation, we can reframe states as goal states. A goal state (or "solution") can be defined as a state of the system that minimises free energy. In Bayesian terms, this state is a posterior belief about the state of the world, given the sensory evidence, that is most likely according to the system's generative model. Inference is the process of arriving at this goal state.

C. Precision Weighting in Goals and Preferences

Our preferences and goals, according to the FEP, are by definition those future states we expect the most. Some anticipated future states are assigned higher *precision*, that is, we have a higher degree of confidence or certainty in their occurrence compared to others. These define our goals and preferences - the states we 'value' more. Precision-weighting of different expected outcomes changes over time depending on context. Multiple factors feed into precision-assignments at a given point in time, potentially including the following:

- **Biological need.** In the context of homeostasis, predictions about the physiological needs of the body (like maintaining a certain body temperature, or energy level) are likely to be assigned high precision, reflecting the central value of these needs. When these needs are not met (or expected not to be met), the body experiences a large prediction error (or expected prediction error). This motivates policy selection to reduce this error such as getting somewhere warm or finding food. In this case the precision of our expectations of those states is hard-wired in our biology.
- **Habit.** Familiar or habitually experienced states and associated policies over time gain high precision because they've been confirmed repeatedly in the past and confidence in them occurring again is high. When these states are not experienced there is a large prediction error (or expected prediction error), which motivates policies to reduce this error by selecting policies that habit. In this case the precision of our expectations of those states is based on the accumulated evidence (sensory data) associated with those states.
- **Learning and Experience:** Related to habit-formation, as we gain more knowledge and experience, the precision of our predictions concerning certain states being attained can increase. This reflects the accumulation of evidence in favour of these predictions.
- **Emotional Relevance**: States or outcomes that have strong emotional relevance might be given higher precision. For instance, predictions tied to outcomes associated with high reward/pleasure or high threat/pain will be

assigned high precision, reflecting their salience and importance to the individual.

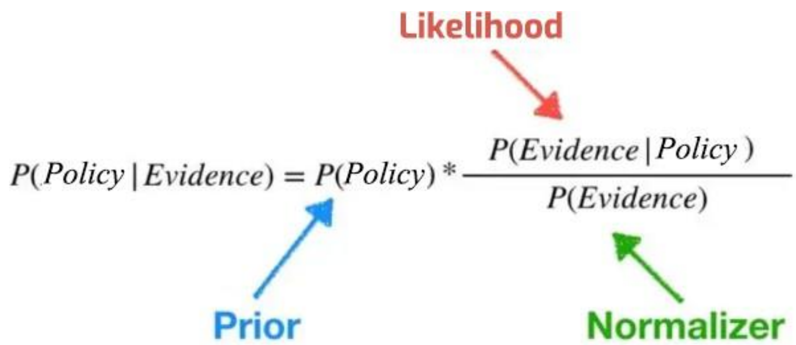
- **Social Context**: Culture and social learning and can influence precision weighting. For example, beliefs or values imparted from parents, peers, or society at large can be assigned higher precision due to their social importance.
- **Attention**: The direction of attention can also modulate precision. When we pay more attention to certain stimuli, the brain often treats the predictions associated with these stimuli as having a higher precision according to the FEP. This is sometimes called ‘attentional gain’.
- **Uncertainty and Variability**: In situations with a high degree of uncertainty or variability, the brain may decrease the precision weighting assigned to predictions related to these situations. This allows the brain to be more flexible and exploratory through *epistemic* active inference, allowing for learning and updating of the internal model when new information is received. Here the prioritised goal is ‘gaining more relevant information to reduce uncertainty’.

These examples show how precision-weighting plays a crucial role in defining and prioritising our goals or preferences by influencing the confidence or certainty we have in certain future states or outcomes. Higher precision-weighting leads to a higher degree of confidence in a certain prediction, effectively determining how much we value or 'care' about a prediction error related to that prediction.

D. Choice & Bayesian Decision-Making

The higher the precision-weighting of a predicted outcome, the more surprise we will experience if the outcome does not occur and the stronger the motivation to implement policies to reduce the resulting expected free energy.

There is a parallel between the inference process in selecting policies in active inference and belief selection in perceptual inference.



Here the likelihood term P(Evidence|Policy) can be interpreted as the probability of observing the sensory data (or states) conditional on what we would expect to see as a result of executing a certain policy - our planned sequence of actions.

By multiplying the prior (probability of a policy based on previous experience) with the likelihood, we compute a posterior probability that represents the updated belief in the efficacy of a policy given current evidence. In the same way that a mismatch between predicted and actual sensory data drives perceptual updating in Bayesian inference, a mismatch between the sensory outcomes we expect from a certain policy and the sensory outcomes we actually experience can drive policy updating in active inference. If a policy doesn't yield the expected sensory outcomes (i.e., doesn't get us closer to our goal or result in the reward we expect), this can be thought of as a prediction error - an indication that our current policy needs revision. By extension, if we are using Bayes inference to evaluate between multiple policies, actions are selected according to the policy that maximises the posterior belief (minimises free energy).

As with Bayesian perceptual inference, precision-weighting of the policy (prior) and likelihood (policy evidence) also plays a critical role in policy selection.

When selecting between possible courses of action, we take into account not only which action is expected to bring us closest to our goal or goals, but also our confidence in those policies. High precision policies for which we have accumulated a lot of evidence will be selected over low-precision policies where there is more uncertainty or variance in outcomes. And when the states we experience as a result of our action plans are clear and unambiguous, we will give more weight to this evidence in updating our policies. The balance between these two sources of precision weighting also influences the trade-off between *exploitation* (sticking with familiar, reliable strategies and goals) and *exploration* (seeking out new information and experiences to update both goals and strategies).

In summary, Bayesian inference is used as a model for both perception and action according to the FEP, reflecting a deep symmetry between perception and action.

E. Time Horizons and Cost-Benefit Tradeoffs

On the FEP, our chosen policies to realise preferences and goals are ultimately determined by expected free energy minimisation.

The expected free energy under a particular policy is calculated as a sum of the free energy expected for each time step within the future planning horizon. When evaluating policies, take into account the free energy associated with all future states and actions predicted under that policy. So, the selection of policies is based on a full trajectory of states and actions, and the aim is to minimise the total or cumulative free energy across this trajectory. In this way, decisions are informed by considering the longer-term consequences of chosen actions, not just the immediate outcomes.

This underpins *short-term vs long-term cost-benefit tradeoffs.* While policies that are associated with a lower cost (time, effort, resource use, negative emotions) and risk (harm, loss) for a valued goal are generally prioritised, often long-term expected free energy reduction will be greater combined with short term costs or risks.

For example, in Walter Mischel' classic Marshmallow Test, a child is given a choice: they can have one marshmallow now, or if they can wait for 15 minutes they will be given two marshmallows. The first policy is to eat the marshmallow now, which will result in an immediate reward (sensory pleasure) and will minimise immediate prediction error (the expected sensory pleasure of eating the marshmallow). The second policy is to wait and get two marshmallows later, which involves delaying the reward and enduring the cost (discomfort, frustration) of waiting.

The expected free energy for each policy is computed based on the sensory outcomes associated with each action, as well as the predicted or preferred sensory states. For the first policy, the expected free energy is low in the short term, because the sensory prediction (pleasure from eating a marshmallow) matches the actual sensory outcome. However, over the longer term, this policy may be associated with higher free energy, because the child would experience disappointment or regret from not getting the second marshmallow.

On the other hand, the second policy (waiting) might be associated with higher expected free energy in the short term, because of the prediction error associated with not eating the marshmallow right away. However, over the longer term, this policy might lead to lower free energy, as the child gets to enjoy two marshmallows and the pleasure of achieving their goal.

According to the FEP, the child's decision would be guided by the policy that minimises expected free energy over the *planning horizon*. If the child is able to anticipate the future reward (and tolerate the immediate discomfort) they should choose the second policy. So this decision-making process involves a tradeoff between immediate and future expected free energy.

F. Simplicity-Complexity Tradeoff: Occam's Razor

The Kullback-Leibler (KL) divergence term in the free energy formula (described above) represents the cost of maintaining more complex models of the world. This complexity cost term in the FEP introduces an important extension to the traditional 'Bayesian brain' framework.

A complex model can potentially provide a more accurate representation of the world, allowing for more precise predictions resulting in a reduction in surprise and prediction error. However, we need models that are not only accurate but also generalisable, interpretable, robust and internally consistent - principles that benefit from simplicity:

- **Generalisation**: Simpler models are indeed often more generalisable. Overly complex models that fit every detail of the training data can lead to overfitting, where the model performs well on the training data but poorly on unseen or future data. Simplicity helps in capturing the essential features that are likely to generalise to other contexts.
- **Abstraction**: Simplicity in models often corresponds to a higher level of abstraction. By focusing on the essential features and ignoring unnecessary details, simpler models can recognize common patterns and relationships across different contexts. This abstraction is indeed vital for general

intelligence, as it allows for the application of learned knowledge in various situations.

- **Robustness**: Simpler models are typically more robust to noise, especially when there is uncertainty in distinguishing between signal and noise. Complex models may inadvertently fit the noise in the data, mistaking it for a genuine pattern. Simpler models, by focusing on the essential features, are less likely to be swayed by random fluctuations or noise in the data.
- **Internal consistency**: Complex models often have more parameters, relationships, and dependencies. This increased complexity can lead to more internal inconsistencies through conflicting assumptions, contradictory relationships, or misalignment with sensory experience. And these inconsistencies are more difficult to detect and resolve when models are more complex.

Moreover, a more complex model requires more resources to build, maintain, and update. These resources include energy, time, and cognitive processing capacity. Therefore, from the standpoint of internal consistency and resource efficiency, a simpler model may be preferable. An optimal internal model would have just the right level of complexity needed to accurately represent the world and make reliable predictions, while also being as simple as possible to conserve resources, adopting Occam's razor .

The need for Occam's razor can be extended conceptually to policies in active inference. Policies that involve a series of complex and uncertain actions may lead to a higher overall divergence, and thus prediction error, due to the increased uncertainty associated with their outcomes. The more complex the policy, the more potential there is for unexpected outcomes, and thus the greater the divergence from the agent's predictions under the policy.

G. Active Inference: Proactive & Reactive Control

One way of interpreting the distinction between free energy (*F*) and expected free energy E(*F*) is that the inference to minimise *F* comes *after* the free energy is computed in the case of *F*, and *before* the free energy is computed in the case of E(*F*).

Free energy minimisation (*F*) under the FEP can be seen as a kind of *reactive control* process because it involves minimising discrepancies between our sensory input and our predictions about that input in the present moment: it involves responding to the present evidence to reduce surprise.

Expected free energy minimisation, on the other hand, is a type of *proactive control* because it is about minimising the anticipated or expected surprise over future time points. This is often associated with active inference or selecting actions (policies) that are expected to lead to future sensory states that minimise surprise. In this sense, it involves anticipating and acting to shape the future evidence to reduce surprise.

In terms of the traditional distinction between reactive and proactive control in cognitive neuroscience, the former is reactive, and the latter is proactive.

5. FEP & The Life Sciences

According to the FEP, not only humans but all life forms are self-organised to minimise surprisal in the environmental niche to maintain structural and functional integrity.

A. Informational and Thermodynamic Entropy Minimisation

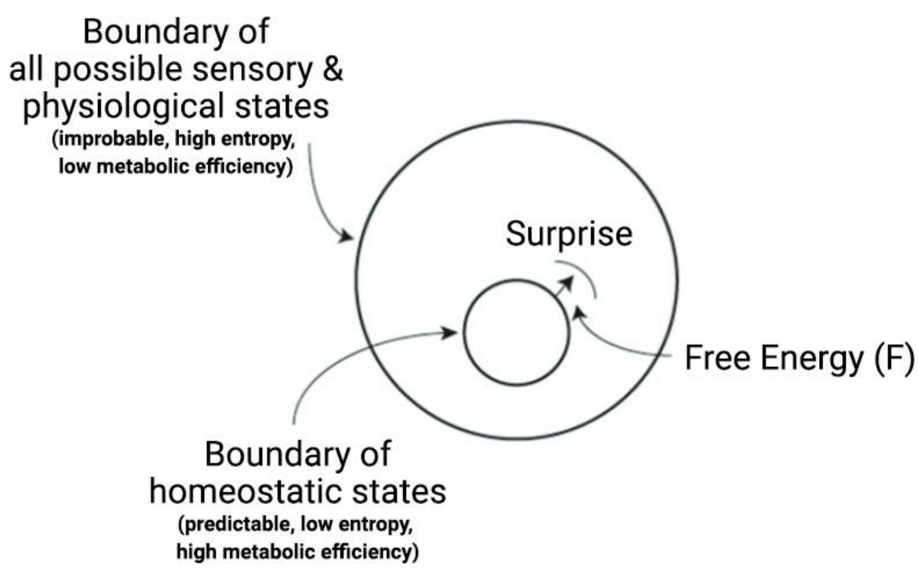
Surprisal is an information-theoretic measure of the information content associated with an event or outcome. Here we will use the term 'surprise' for surprisal. 'Information' refers to a reduction in uncertainty about the state of the world. Events that are less predictable or more unusual have greater surprise, as they convey more information than events that are more predictable or occur frequently. Information *entropy* is a measure of the average surprise. High entropy implies high uncertainty because outcomes are very unpredictable and therefore surprising on average. Low entropy implies low uncertainty because outcomes are very predictable and therefore not very surprising on average.

There is a deep connection between information entropy and *thermodynamic entropy* that measures the degree of disorder or randomness in a physical system. Both concepts of entropy have to do with uncertainty, unpredictability, and the dispersion of a distribution - whether a distribution of microstates of a physical system (in the case of thermodynamic entropy) or a probability distribution over outcomes (in the case of information entropy). Moreover, if you assume that all microstates in a physical system are equally probable, the formula for thermodynamic entropy is identical to the formula for information entropy.

B. Homeostasis & Metabolic Efficiency

The FEP bridges information theory and statistical physics by proposing that organisms resist thermodynamic entropy by minimising free energy (*F*). Minimising F reduces not only information entropy, but also physiological (thermodynamic) entropy, since it maintains the organism in a limited set of high-probability, low-entropy states. These states can be understood in terms of *homeostasis* - that is, the maintenance of a constant internal environment in the body despite changes in the external environment or the body's own activities which allow the cells, tissues and organs, and organism to function at optimal metabolic efficiency. For example, the body maintains an internal temperature of approximately 37°C because the enzymes that catalyse our metabolic reactions work most efficiently at this temperature. If our body temperature drops or rises too much, these enzymes will not function as effectively, and metabolic processes are impaired.

This is a way of understanding homeostasis, as shown in the figure below.



To use Friston's example, a fish out of water would be in a surprising state and a fish that frequently was out of water would have high entropy, far from its homeostatic ranges that preserve its functional and structural integrity and metabolic efficiency. Biological organisms in general must therefore minimise *F* and the long-term average of surprise to ensure that they maintain homeostasis and metabolic efficiency with both their sensory (informational) and physiological (thermodynamic) entropy remaining low according to the FEP.

C. Adaptation & Evolution

On the Free Energy Principle, over ontogenetic (lifespan) and phylogenetic (evolutionary) time scales, organisms become models of their cognitive-environmental niche through the internal organisation of physiological systems, nervous systems, DNA and genetic regulatory networks. These models allow them to predict and adapt to their environment. For example, an animal with memory-networks might learn to predict when and where food will be available (external states) based on sensory observations (the time of day, seasonal temperature, etc.). If this model is accurate it reduces the uncertainty or surprise concerning finding food, and by reducing uncertainty and surprise can more effectively maintain its energy balance and continue to survive and reproduce. At the same time, these brain-based models drive behaviour that actively changes the environment to make it more predictable - for example by caching food via active inference, which in turn promotes adaptive fitness.

In this way, over different evolutionary and developmental timescales, organisms embody their own generative models and cognition is fundamentally grounded in the organism's adaptive interactions with its environment.

To summarise some key ideas:

- **Minimisation of Free Energy as an Evolutionary Imperative**: On the FEP living organisms strive to minimise free energy or reduce uncertainty in their sensory inputs via generative models. These may be embodied in DNA and genetic regulatory networks, which underpins natural selection, where organisms with traits that allow them to survive and reproduce in their environments are more likely to pass on their genes to future generations.

- **Prediction and Selection**: The FEP involves the use of internal models to predict sensory inputs and guide action. These models are continuously updated based on prediction errors or surprise, in a way that could be seen as analogous to natural selection: less accurate models (analogous to less fit organisms) are discarded or updated, while more accurate models (analogous to more fit organisms) are kept and used to guide future predictions and actions.

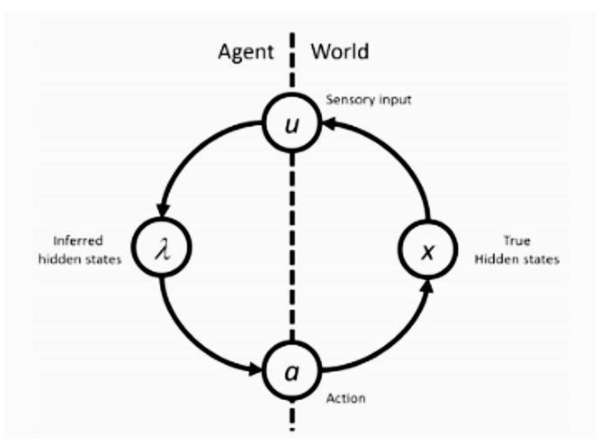
- **Exploration - Exploitation Trade-off**: According to the FEP, organisms must balance exploitation (using existing models to minimise free energy) and exploration (updating or generating new models in the face of surprise). This aligns with natural selection's balancing act between preserving successful genetic variations and generating new variations for potential adaptation to changing environments.

D. Internal-External Boundaries, Nesting & Autopoiesis

Organisms are dissipative systems far from thermodynamic equilibrium and require a continuous input of energy to maintain their structure and function. They are also autopoietic in the sense that they are autonomous, with sufficient internal processes to produce and maintain themselves in structure and function. All such systems maintain a boundary that separates their internal states from their external milieu.

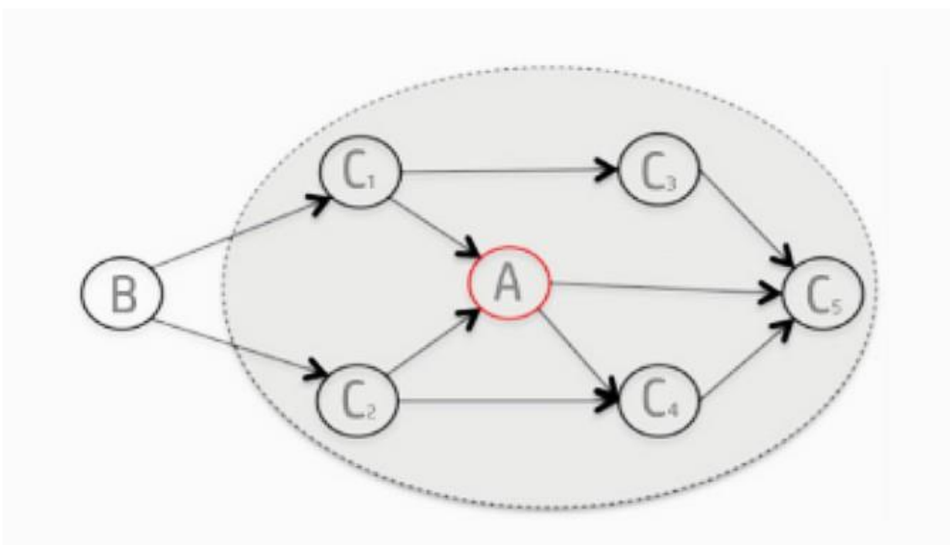
> *"A cell stands out of a molecular soup by creating the boundaries that set it apart from that which it is not. Metabolic processes within the cell determine these boundaries. In this way the cell emerges as a figure out of a chemical background. Should this process of self-production be interrupted, the cellular components … gradually diffuse back into a molecular soup."* Varela, Maturana & Uribe

According to the FEP, the boundary consists of sensory and active states shown here:



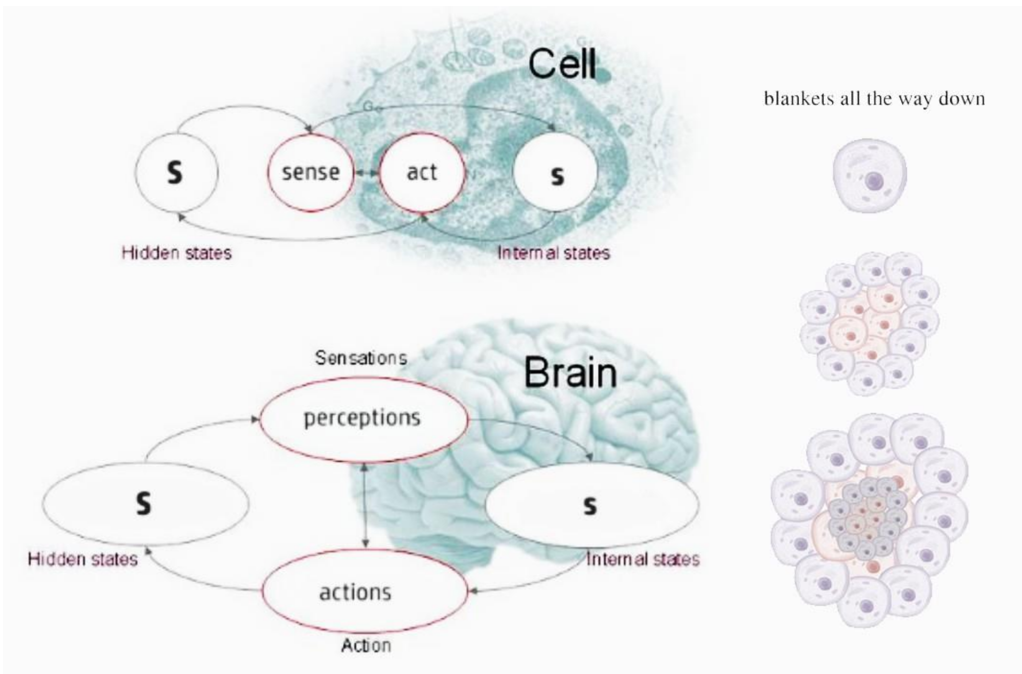
The FEP posits that *Markov blankets* define the boundaries of such dissipative systems. A Markov blanket is a concept from the field of probability theory and statistics and is used in Bayesian networks (a type of statistical model) to describe the set of variables that shield or insulate a given variable from the rest of the network. Extending on the ideas discussed above, according to the FEP, it is Markov blankets that keep dissipative systems removed from thermodynamic equilibrium.

The existence of a Markov blanket means external states are (in a statistical sense) conditionally independent of internal states, and vice versa. Internal and external states can only influence each other via sensory and active states. A is conditionally independent of B given C if, when C is known, knowing A provides no further information about B. This maps onto the Markov blanket (C1-C5) for node A shown here:



Once all the neighbouring variables for A are known, knowing the state of B provides no additional information about the state of A. According to the FEP, the internal, blanketed state (A) constitutes the model. The 'children' of the model (C4, C5) are the *active states* that drive action through prediction error minimisation in active inference. The sensory states are the 'parents' of the model (C1, C2), driving *perceptual inference*. If the hidden causes (B) beyond the blanket are inferred accurately, the system minimises free energy. External states (B) which are 'hidden' beyond the Markov blanket thus cause sensory states (C1, C2), which influence, but are not themselves influenced by, internal states (A), while internal states (A) cause active states (C4, C5), which influence, but are not themselves influenced by, external states.

Autonomous living systems tend to comprise not only the unified organism itself, but a multiplicity of nested systems, each of which models its external world adaptively. These can be realised by multiple hierarchically nested Markov blankets. Thus a living system can be understood as composed of Markov blankets of Markov blankets — reaching all the way down to cellular organelles and DNA and all the way up to whole brain networks.



Adapted from Friston, 2013

And the boundaries of such systems need not in principle be constrained by the biological boundaries of a living organism. In principle, the FEP may be applied if ‘internal’ states are shared across interacting individuals with collective intentionality - such as with a flock of birds or a human cultural group with shared cultural infrastructure, goals and values. The relevant level of analysis in evolutionary theory

Global free energy minimisation on this account involves multiple nested generative models in the *hierarchical network*, each accountable to the others, providing an internally consistent representation of sensory causes at multiple levels of abstraction over *multiple time scales*. This gives generative models in complex organisms spatiotemporal depth, enabling the overall biological system to make inference over recursively larger and larger scales of sensorimotor consequences.

E. Allostasis & Adaptive Autonomy

Biological systems are homeostatic systems with dependencies over multiple time scales. They are able to actively monitor and *react* to perturbations that challenge homeostatic variable ranges which may go out of bounds by minimising the resulting free energy - such as shivering when temperature falls. They are also able to *predict* challenges to homeostatic variables in advance to maintain homeostasis, and implement strategies to minimise expected free energy through active inference. The term *allostasis* refers to the process by which the body maintains stability (homeostasis) by anticipating and adapting to changes in the environment before they occur. Allostasis can be considered as a kind of 'predictive regulation' of the body's internal environment, through expected free energy minimisation. It's about preparing for the future, not just reacting to the present as in negative feedback based homeostasis (also see section 4F on proactive and reactive active inference control above).

Thus biological systems possess generative models with temporal depth, sampling among different options and selecting the option that has the least (expected) free energy or greatest expected evidence. In this way, living systems are able to ‘free’ themselves from reacting to proximal conditions by making inferences about probabilistic future states over multiple scales. According to the FEP, this kind of adaptive, future-directed active inference is what grounds an organism’s autonomy. Living systems can transcend their immediate present state, and work towards occupying future states with an expected free energy minimum.

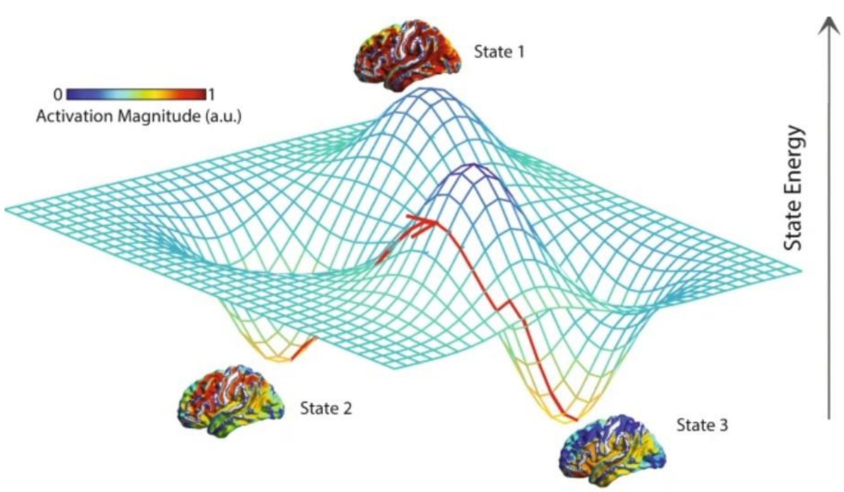
And according to the FEP, adaptive autonomy depends on hierarchical generative models with nested and multi-layered Markov blankets (section 5C). It is only with multilayered independencies that living organisms can learn (within a lifetime) and evolve (over multiple generations) generative models with temporal and spatial depth, enabling the system to make inference over larger and larger scales of sensorimotor consequences (Kirchoff et al., 2018).

III. Brain Energy Landscapes

The Free Energy Principle (FEP) has been extended and generalised in multiple ways since its original formulation by Karl Friston in 2005. Below we review one constellation of potential extensions.

The concept of an energy landscape is used to describe the different network states that the brain can be in, and how 'energetically attractive' these states are. The most probable states are those that the brain spontaneously adopts because they require less energy to maintain. These are lower energy brain states in the sense that they are less metabolically costly, requiring less oxygen and glucose consumption. In the context of the Free Energy Principle, the brain is driven to minimise free energy (F). By minimising F, the brain is effectively navigating the landscape of possible states to find those that are probable and low energy.

This concept is related to the idea of an attractor landscape in dynamical systems theory. Under extended FEP, the 'hills' in such a landscape represent higher energy, low probability states (repellors) while the 'valleys' (attractors) are states that are highly probable and low energy. The brain is driven to seek out these low-energy states to minimise surprise or prediction error. In this case, the depth of the valleys could be thought of as inversely related to their free energy and metabolic energy cost. This kind of landscape is shown here:



Gu et al., 2018

The extended FEP posits that the brain is inherently driven to navigate this landscape in a way that minimises free energy and metabolic cost by transitioning from any starting point on higher ground to a high-probability low energy state - an ‘attractor’. In this way, the brain over the long term resists surprise and entropy and maintains homeostasis. These low energy states are considered to be stable configurations of brain activity.

Surprise or prediction error results in more information to process which results in an increased metabolic cost: the neural processing of information is energy consuming. Additionally, there are energetic costs associated with the complexity of our generative models: more complex representations require more energy to maintain them in a resting state. So not only is the brain driven to stay within a range of easily accessible, high probability psychological states that flow from its generative model, but also to reduce the energetic cost of model encoding in neural networks.

The brain can thus be viewed as striving to balance the metabolic cost of neural computation and physiological response with the need to reduce free energy by accurately predicting and responding to its environment, consistent with F being a function of both prediction error and the complexity of the generative model.

Shi Gu and colleagues (2018) have looked at how large-scale brain circuitry constrains states of neuronal activity and transitions between those states, and how diverse cognitive systems are optimised for differential contributions to integrated versus segregated function via distinct patterns of energy utilisation.

They have revealed that the sensorimotor networks and networks involved in complex cognitive functions such as attention, working memory, and cognitive control - the fronto-parietal, salience, dorsal attention, and cingulo-opercular network - are associated with more energetically costly states in the brain's energy landscape. These networks, when actively engaged in processing complex information, require more energy to function. These information-processing states are less probable, meaning they don't occur as frequently as the more energy-efficient states associated with the default mode network (DMN) that is active during rest.

Gu and colleagues highlight that (i) energy consumption and production varies by brain region, (ii) sets of regions that show strong coherence in cognitive systems require high amounts of oxygen and glucose consumption, (iii) regions with high levels of functional connectivity consume high levels of energy (glucose) and require greater blood supply.

When the frontoparietal and attention networks are in an efficient state with minimal energy, they are more internally focused and have low between-system energy; that is, they are not expending much energy in communicating with other networks. They may be considered functionally isolated during low-energy states. There are, however, situations where these networks - known for their network hub status - have a high level of energetically costly communication with other networks. They are part of higher-order cognitive systems that display flexible, adaptive activation rates, and have the capacity to be extensively connected to other networks during complex cognitive processing.

The sensorimotor system, along with primary and secondary sensorimotor cortices in somatosensory, visual, and auditory systems is part of a cluster with high between-system energies when these are engaged in tasks that require working memory, attention and cognitive control. In such low probability states, the brain shifts to state configurations with higher energy expenditure due to the increased complexity of neural computations and coordination required. Conversely, when the sensorimotor cluster is engaged in more automatic or habitual tasks, they operate with relatively lower energy expenditure, possibly in conjunction with the default mode network (DMN) that is active at rest.

The DMN has intermediate between-system energies. The DMN is generally associated with a lower energy landscape because it is more active during resting states, which are energetically more efficient for the brain. Within these lower energy states, the DMN has intermediate between-system energies, reflecting its balanced role in internal processing and interaction with other networks.

To summarise some key concepts reviewed above:

A. Attractor States

The idea that the brain seeks out low-energy, high-probability states (attractors) is central to the energy landscape model. The brain's overall goal is to reach a state of minimum free energy, where surprise or prediction error is reduced. This is an energy-efficient state, and it's in these states that the brain may be understood to do most of its information processing work.

B. Metabolic Cost Tradeoff

The balance between reducing free energy and managing metabolic costs is an important one. Neural computation, physiological response, and maintaining the complexity of our generative models all require energy. The brain is always looking to strike a balance between achieving a low-energy, high-probability state in terms of free energy reduction and the energy costs of getting there. This underpins the complexity (accuracy) - simplicity (efficiency) tradeoff in the Free Energy Principle.

C. Functional Connectivity

Functional connectivity refers to the statistical interdependencies or correlations between different regions of the brain. These connections can be mapped out in networks of areas that work together and different cognitive tasks and states (including rest) activate different networks within the brain.

In terms of the energy landscape picture, each of these functional networks can be seen as located in a certain 'valley' or 'peak' on the landscape, depending on how much energy they require to be maintained and how frequently they are activated. Complex cognitive tasks that require top-down, effortful cognitive control, such as decision making and problem solving engage networks like the frontoparietal network (FPN) and cingulo-opercular network (CON). During these tasks, this network processes more information (and thus free energy) and coordinates activity among numerous interconnected regions to process it, which increases the overall metabolic cost in glucose & oxygen consumption. In the energy landscape analogy, these active states would be represented as peaks or high-energy areas.

Conversely, the default mode network (DMN) is more active during resting states and spontaneous, non-directed cognition, and is associated with lower-energy 'valley' states on the energy landscape which are less metabolically costly.

D. Energy Landscapes and Cognition

The energy landscape framework provides a useful perspective on brain function. It links physiological aspects such as metabolism and oxygen and glucose consumption, with cognitive processes and states. Different network states, and transitions between them, represent an ongoing balance between energy conservation and the need to process and respond to free energy and information which is metabolically costly.

IV. Extended Evolutionary Synthesis (EES)

Adaptive evolution involves selection operating on heritable variation resulting in changing frequencies of variants in a population over time. The classical Modern Synthesis (MS) focuses on genetic variation and natural selection resulting in changes in gene frequencies in a population over time, while the Extended Evolutionary Synthesis (EES) expands this view to include non-genetic forms of heritable variation (like epigenetic, cultural, and ecological inheritance) and additional selection mechanisms (like developmental bias and niche construction). The EES was a response to advances in developmental biology, genomics and ecology.

Drawing from a review by Kevin Laland and colleagues (2015), mechanisms of the Extended Evolutionary Synthesis (EES) and how they compare to the classical Modern Synthesis (MS) are outlined below.

A. Generation of Novel Variants (Phenotypic & Genotypic Diversity)

Modern Synthesis: Random Variants

- **Random genetic variation**. The Modern Synthesis proposes that random processes of mutation and recombination are the primary sources of the novel genetic variation that natural selection acts upon. Changes in an organism's genetic material do not have any inherent directionality towards increasing the organism's fitness. Example: A random mutation might occur in a gene that affects human hair colour. This mutation is a random event, not influenced by whether a change in hair colour might make the individual more or less attractive to potential partners. The new hair colour is a novel variant that can then be acted upon by natural selection.

ESS: Directed Variants

- **Developmental processes.** These include epigenetic effects, regulation of gene expression, and construction of internal (hormone levels, nutrient availability, etc) and external (temperature, light levels, food availability, etc) developmental environments. Example: During human development, the nutrition a foetus receives in the womb can influence its growth and development, leading to phenotypic variation.
- **Developmental bias.** Developmental systems facilitate well-integrated, functional phenotypic responses to mutation or environmental induction rather than random ones. Example: The brain's development is tightly regulated by a host of genetic and

environmental factors, and as a result not all changes to the brain's structure and function are equally likely to occur. A mutation that slightly alters the timing of synaptic pruning might lead to subtle changes in brain function, while a mutation that disrupts the existing developmental process is likely to be harmful and therefore unlikely to persist in a population.

- **Niche construction.** Organisms modify their own and each other's niches, which can affect both the generation of novel variants. Example: The development of agriculture led to changes in diet that are thought to have driven the evolution of genetic adaptations to digesting certain types of food, such as lactose tolerance in populations that domesticated dairy animals.
- **Environmental induction**. Changes in the environment can lead to specific phenotypic responses in organisms, generating non-random variation. This includes phenotypic plasticity, where a single genotype can produce different phenotypes depending on the environment. Example: A high-altitude environment induces specific phenotypic responses in humans, such as increased red blood cell production and changes in haemoglobin structure, to cope with low oxygen levels. This is a form of phenotypic plasticity, where a single genotype can produce different phenotypes depending on the environment. Over many generations, these kinds of environmentally induced changes can lead to genetic changes if individuals with genotypes that allow for more effective phenotypic responses to high altitudes have higher survival or reproductive success. This process is known as *genetic accommodation*. In this way, phenotypic plasticity can guide and shape the course of genetic evolution.

FEP & Variants

The FEP posits that adaptive, biological systems minimise surprise or prediction error (free energy) through the continual updating of its internal models to better predict sensory inputs. This process involves an interplay between exploiting known, predictable states (maintaining existing models) and exploring new, uncertain states (updating or generating new models) - a trade-off known as exploitation vs. exploration.

In terms of generating novel variants, the FEP suggests that the brain will sometimes favour exploration over exploitation for long-term free energy minimisation. In cognitive terms, this could be the generation of novel ideas or creative problem-solving approaches that provide new ways of interpreting and interacting with the world. These novel cognitive or symbolic variants then lead to changes in behaviour or modifications of the environment (EES's niche construction) that impact an individual or groups evolutionary trajectory - akin to EES's environmental induction.

In addition, just as the EES recognises the role of environmental and developmental factors in generating phenotypic variation, the FEP recognizes that changes in the environment and an organism's developmental history will channel generative models, leading to variation in its behaviour, phenotype and ecology.

Like the EES, the FEP emphasises the importance of not just genetic, but also epigenetic, phenotypic, and environmental factors in generating variation which then contributes to the. energy landscapes that organisms navigate in their pursuit of minimising long-term free energy and maximising evidence.

B. Selection Mechanisms

Modern Synthesis: One-Way Causation

- **Natural selection.** This is a unidirectional process where the environment shapes the organism, but not vice versa. Traits that enhance survival and reproduction become more common in successive generations of a population, driven by the differential reproductive success of individuals with different phenotypes. Example: The evolution of lactose tolerance in some human populations is often cited as an example of natural selection. In populations that historically relied on dairy farming, individuals who could digest lactose into adulthood had a nutritional advantage and were more likely to survive and reproduce, leading to an increase in the frequency of lactose tolerance in these populations.

EES: Reciprocal Causation

Organisms both shape and are shaped by their environments, creating a feedback loop that can drive evolution. This overarching principle of two-way causation includes mechanisms such as:

- **Developmental bias.** Some phenotypic changes are more likely to occur than others due to the way organisms develop. This can influence the direction and pace of evolution. Example: Human brain development shows evidence of developmental bias as increases in brain size and complexity have occurred repeatedly in human evolution, suggesting that our developmental processes make these changes more likely.

- **Niche construction.** Organisms modify their own and each other's environments, which can affect the selection pressures they face and thus influence evolution. Example: Humans create complex social structures, technology, and modifications to the environment that greatly influences the selection pressures we face.
- **Levels of selection.** The ESS adopts an organism-centred perspective rather than a gene-centric one, while also recognising that selection can occur at multiple levels of biological organisation, from genes to individuals to groups or populations, such as a cultural niche shared by multiple individuals. Example: Cultural practices shared by a group of humans can influence the survival and reproduction of the group as a whole, potentially leading to the evolution of new cultural norms or behaviours.

FEP & Selection

We have reviewed an FEP interpretation of natural selection above.

In the FEP framework, organisms modify their environments to reduce free energy or surprise, aligning with the concept of niche construction in evolutionary biology, where organisms alter their environments in ways that can affect natural selection pressures.

And in line with the EES's assertion of organisms moulding and being moulded by their environment (reciprocal causation), the FEP emphasises the role of active inference in free energy minimisation where organisms act to align their environments with their internal models or predictions, thereby reducing surprise.

Within the energy landscapes model, organisms will also modify their environments (niche construction) to forge energetically efficient paths, consistent with probability landscape conceptions.

C. Inheritance Mechanisms

Modern Synthesis

- **Genetic inheritance**. Genetic inheritance: This is the process by which traits are passed from parents to offspring through genes - the unit of selection. According to the Modern Synthesis, genes are the only mechanism of inheritance, and traits acquired during an organism's lifetime are not inherited. **Example**: Eye colour in humans is primarily determined by genetics. An individual inherits genes from their parents that determine their eye colour.

EES & Inheritance

- **Inclusive inheritance**: This is the idea that inheritance encompasses more than just genes. According to the EES, traits can also be passed on through mechanisms like epigenetic changes (modifications to gene expression), behavioural and cultural transmission (learning from others), and ecological inheritance (changes to the environment caused by previous generations). These can in turn shape genetic evolution by either of the following mechanisms:

    - **Shaping selection pressures**: For example, the development of agriculture led to changes in diet that could have influenced the evolution of genes related to digestion and nutrient absorption.
    - **Cultural transmission can lead to genetic changes over time**: This is a process known as gene-culture coevolution. For example, in societies where reading and writing are important skills, there could be selection for genetic traits that facilitate these abilities, such as better visual acuity or fine motor control, thus changing the frequency of certain genetic traits over time.

FEP & Inheritance

Non-genetic forms of inheritance, such as epigenetic, cultural, and ecological inheritance, have a strong role in the EES in addition to genetic inheritance.

According to the FEP, an organism is constantly learning and updating its model of the world, which includes inherited knowledge. If this information is passed down through generations, it can shape the ways in which future generations predict and interact with the environment, whether it's in the form of learned behaviours, ideas, cultural norms or technological tools.

When viewed in terms of an energy landscape, these inherited traits influence the organism's position and the possible paths it can take. Whether these traits are genetic, epigenetic, behavioural or cultural, they all contribute to the energy landscape and can play a role in shaping an organism's evolutionary trajectory.

D. Macroevolution

Modern Synthesis

- **Gradualism**. Macroevolution is largely a result of gradual changes that accumulate over long periods of time. This is often referred to as gradualism, in which phenotypic transitions occur through multiple small steps, leading to gradual evolutionary change. This perspective is based on the idea that evolution via mutations of large effects is unlikely because such mutations have disruptive pleiotropic effects - that is, maladaptive effects on multiple unrelated traits.

EES & Macroevolution

- **Variable rates of change.** The ESS posits that variants of large effect are possible, which can lead to rapid evolutionary change. This theory allows for ‘saltation’ or ‘punctuated equilibrium,’ where significant evolutionary changes can occur rapidly, via either of the following mechanisms:

    - **Mutations in major regulatory control genes** expressed in compartment-, tissue-, or module-specific manner. Compartment- specific expression might refer to a regulatory gene that is active in a specific compartment of a cell, such as the nucleus, where it controls the expression of genes within that compartment. Tissue-specific expression might refer to a regulatory gene that is only active in heart tissue. Module-specific expression could refer to a regulatory gene that is active in a specific module of the brain, such as a functional network.
    - **Developmental processes** respond to environmental challenges with changes in coordinated suites of traits, or through nonlinear threshold effects where a small change in an environmental factor or a genetic change can lead to a sudden and significant change in an organism's phenotype or behaviour. These developmental processes could be responses to selective pressures from niche construction and ecological inheritance.

FEP & Macro-Evolution

EES allows for sudden shifts in evolutionary rates and accepts the hypothesis of ‘punctuated equilibrium,' wherein marked evolutionary changes can occur quickly.

The extended FEP interprets this as the organism rapidly changing its predictive models - either in response to drastic changes in sensory input or through internal re-organisation of generative models. .In energy landscape terms, these sudden changes can signify an organism moving rapidly to a new ‘attractor’ energy minimum, potentially due to significant environmental alterations or major mutations.

E. Multiplier Effects & The EES

The gene-environment multiplier effect (Dickens and Flynn, 2001) describes how genes and environment can interact in a way that amplifies the effect of each on the phenotype of an individual such as cognitive ability. An individual with a small genetic advantage in intelligence can influence the environments they seek out or create for themselves, such as advanced educational opportunities or intellectually challenging jobs which in turn enhance their cognitive abilities. Enriched environments may also upregulate genes associated with cognitive function, further enhancing the person's intellectual abilities. These interactions can amplify over time. Small initial differences in genetic predisposition can lead to increasingly divergent environments, which in turn lead to increasingly divergent gene expression and phenotypic outcomes. This process can lead to a *positive feedback loop*, where genetic predispositions and environmental influences mutually reinforce each other, leading to substantial phenotypic differences over time. Multiplier effects align with the EES in the following ways:

- **Phenotypic Plasticity and Environmental Induction**: A single genotype can produce different phenotypes depending on the environment. This is a form of environmental induction, where changes in the environment - such as intellectual enrichment - can lead to specific phenotypic expressions in individuals, such as stronger cognitive abilities.
- **Reciprocal Causation**: The notion that individuals can seek out environments that are favourable for their intellectual development is an example of reciprocal causation, where organisms not only are shaped by their environments, but also shape their environments.
- **Inclusive Inheritance**: The multiplier effects resulting from small modifications in biological and social environments is a type of non-genetic, cultural inheritance. If these modifications lead to changes in behaviour or cognitive abilities that are then passed on to subsequent generations (through learning or social influence), this could contribute to evolutionary change.
- **Niche Construction**: The process of individuals seeking out and creating favourable environments for their intellectual development is a form of niche construction, where organisms modify their own and each other's environments, which can in turn affect the selection pressures they face.

FEP & Multiplier Effects

The gene-environment multiplier effect — the mutual reinforcement of genetic predispositions and environmental influences — is compatible with the FEP through active inference and niche construction.

Multiplier effects in extended FEP terms, could be understood as driving increasing capacity for inference and complexity and accuracy of generative models, and expanding the space of metabolic energy-minimising trajectories through the organism's energy landscape - extending out to its environmental niche.

F. Summary

In conclusion, the EES, FEP, and energy landscape concept share a common emphasis on the reciprocal interactions between organisms and their environments in determining evolution and behaviour. They all underline the importance of non-genetic forms of inheritance and the capacity of organisms to modify their environments, as well as the influence of environmental alterations on the development and evolution of organisms. These frameworks complement and reinforce each other in their understanding of biological systems, providing a more nuanced and holistic view of the dynamics of life.

V. Complex System Entropy & Syntropy

The concepts of Entropy and Syntropy (or Negentropy) are found in various fields of research including physics, information theory, complex systems and network theory, ecology, cognitive neuroscience and AI. Here we use an interdisciplinary approach to situate the constructs in complex system extensions of the FEP.

A. Entropy (*H*) in Information Theory

Flexibility Under Uncertainty

In information theory, entropy is a measure of the unpredictability or randomness of a system. In extensions of the FEP, this uncertainty could provide the cognitive 'space' for the agent to explore novel strategies and solutions, enhancing adaptability.

Resource Allocation

Just as entropy in information theory quantifies the ‘space’ available for encoding information, *H* in an extended FEP represents the cognitive space or capacity available for adaptive responses. This cognitive space allows for the exploration of new strategies and solutions.

Redundancy and Robustness

In information theory, entropy can also relate to the redundancy in a system. Similarly, a high *H* in an extended FEP model can imply a level of cognitive redundancy that makes the system more robust against errors or unexpected changes.

B. *H* in Thermodynamics & Complex Systems

Energy Dissipation & Efficiency

In thermodynamics, entropy is related to the dispersal of energy. In an extended FEP, a high *H* can be seen as a complex system's ability to dissipate neural network energy costs across low-energy landscape pathways, allowing for metabolic efficiency. A complex system with high *H* can use its available energy resources efficiently for adaptation, similar to how thermodynamic systems with high entropy are often at equilibrium and thus energy-efficient.

C. *H* in the ESS

The concept of entropy in an extended FEP aligns with the EES emphasis on *phenotypic plasticity*, where a single genotype can produce different phenotypes depending on environmental conditions. A high *H* implies an organism has a greater capacity for phenotypic plasticity. For intelligent systems, this adaptability could be a form of cognitive plasticity, where the organism can flexibly adjust its cognitive models in response to environmental changes, akin to how phenotypic plasticity allows for different physical traits.

D. Syntropy (*J*) in Information Theory

Syntropy (negentropy) implies a form of coherence or 'useful information'. In information theory, negentropy measures how far a distribution is from a Gaussian distribution, which is considered the most disordered or random among distributions with the same mean and variance.

Negentropy (*J*) in information theory is a tool or principle that helps you design more effective, better optimised algorithms or models for processing that data. Negentropy can be used for the inference process itself, as well as the encoding of information in the model, and we can map this distinction onto the Free Energy Principle distinction between the inference processes (active or perceptual) and the resulting generative model.

Inference

- **Anomaly Detection**: In statistical inference and machine learning, negentropy can be used to detect anomalies or outliers. This is related to prediction error and is akin to the brain's ability to detect something unusual in the environment, which may require immediate attention or action.
- **Feature Extraction:** Negentropy can help identify non-Gaussian features in the data - i.e. which features in the data are most meaningful. This helps in the inferential process where the brain tries to make sense of sensory data by focusing on the most meaningful features, as it interprets incoming sensory information.
- **Noise Reduction**: Negentropy can be used to distinguish between the useful non-Gaussian signals and Gaussian noise. By focusing on the components with higher negentropy, one can filter out noise and improve the quality of the signal. This could be considered an active inferential process where the brain filters out irrelevant or distracting sensory information to focus on what's important.

Modelling

- **Independent Component Analysis (ICA):** Negentropy is used to separate mixed signals into their independent components. This is an algorithmic optimization to ensure that the separated signals are as independent as possible. In the context of FEP, this could be akin to the generative model that the brain uses to separate and interpret different sources of sensory information.
- **Data Compression**: Negentropy can also be used in data compression algorithms. A distribution with higher negentropy is farther from a Gaussian distribution and thus contains more structure that can be exploited to compress the data more efficiently. This can be seen as part of the generative model where the brain efficiently encodes information for storage and future use. The brain has to balance the complexity and efficiency of its internal models, as with data compression algorithms.
- **Optimal Coding**: In neural networks and machine learning, the concept of negentropy is related to the idea of sparse coding, where the goal is to represent data using the fewest number of active neurons or basis functions. A higher negentropy in the coding implies that the representation captures more essential features of the data.In the context of FEP, the brain uses sparse coding to efficiently represent the causes of sensory inputs. This is about how the brain's model is structured to best explain the sensory data it receives.

A higher capacity for syntropy in a model indicates a greater ability to extract meaningful patterns from sensory data and adapt effectively. While negentropy is a tool for optimisation—making a given process as efficient as possible—syntropy focuses on the capacity for signal processing - the goal of negentropy optimisation.

E. *J* in Thermodynamics & Complex Systems

Syntropy & Energy Cost

Increasing syntropy in a neural system generally requires more metabolic energy to build more complex information processing networks that effectively model the world. There is a thermodynamic cost of maintaining a complex, adaptive system, far from thermodynamic equilibrium. Moreover, complex cognitive tasks that require high levels of syntropy (like problem-solving or planning) often come with a metabolic processing cost. This aligns with the idea that fluid intelligence tasks, which are effortful, also require metabolic expenditure.

Syntropy & Energy Availability

Both syntropy and Gibbs Free Energy can be thought of as resources that a system has at its disposal. Gibbs Free Energy is the useful energy available for doing work in physical systems, while syntropy represents the cognitive resources available for adapting to new situations or solving problems in cognitive systems. A system with more syntropy has more metabolic capacity and can do more cognitive work.

Syntropy & Energy Efficiency

An intelligent system with high syntropy will be more efficient at using its available free energy. In other words, it will require less energy to perform the same amount of work, due to the optimised information processing as described above. The FEP also captures Occam's razor by penalising models that are more complex than necessary. A system with high syntropy would be adept at constructing models that are as simple as possible but as complex as necessary, thereby efficiently using its Gibbs free energy.

In summary, having more syntropy implies a greater store and more efficient use of Gibbs Free Energy (which should be distinguished from the 'free energy' in the FEP. Developing more syntropy in a complex, dissipative system also implies more energy is needed for its development, maintenance and operation, especially during complex cognitive tasks.

F. *J* in the ESS

Syntropy in an extended FEP relates to core concepts of the EES:

- **Constructive Development:** Syntropy aligns well with the EES concept of constructive development, where organisms actively shape their development through interactions with their environment. The ability to anticipate, problem-solve, and adapt to future challenges (syntropy) can be seen as a product of constructive developmental processes.
- **Reciprocal Causation**: Syntropy also fits into the EES principle of reciprocal causation, where proximate causes (like cognitive abilities) can become ultimate causes that influence evolutionary trajectories.
- **Niche construction:** Syntropy involves the ability to anticipate, problem-solve, and adapt for future challenges, which are all essential skills for niche construction. High syntropy supports the creation of technologies, cultural practices, institutions, and systems.

- **Phenotypic evolution**: According to the EES, evolutionary theory is a theory of phenotypic evolution, defined as transgenerational change in the distribution of heritable traits of a population. Syntropy is a high-level information processing phenotype that enhances an organism's ability to generate novel, adaptive responses to environmental challenges, extending to technologies and social systems. Such phenotypes can be inherited.
- **Goal directed deviation from population templates:** Building on the idea of negentropy as a measure of 'distance to normality,' the 'species template' could serve as a baseline or normal population model. Developmental biases, niche construction, multiplier effects, and environmental induction could be seen as forces that push an individual organism's generative model away from this species template, thereby increasing its syntropy. Just as energy is required to create order in physical systems, energy input in the form of food, social interaction, learning, etc., could be required to create the unique traits and behaviours that make an individual's generative model more complex and less 'normal'. In the context of a, developmental biases could shape syntropy, making an individual particularly adept at certain kinds of problem-solving.

Environmental Induction & Niche Construction

The relationship between entropy (*H*) and syntropy (*J*) can be likened to the EES concepts of environmental induction and niche construction. While syntropy drives the ability to construct niches, entropy measures the adaptability gained from these constructed niches. An organism with high syntropy would be adept at niche construction, creating cognitive or physical environments that are conducive to its own adaptability. Once these niches are constructed, an organism with high entropy would be better equipped to exploit these niches for various adaptive purposes. This aligns with the EES concept of organisms shaping their own evolutionary pathways through niche construction. For instance, by using syntropy to construct a wider range of tools and systems in a 'cognitive niche,' the organism gains more opportunities to reduce free energy through perceptual or active inference or to creatively explore new possibilities. *H* then serves as a measure of how well the organism can utilise these constructed niches for adaptability and evolution.

Biassing Causes

In the EES, biassing causes such as developmental bias and niche construction can guide or constrain the direction of evolution. While syntropy (*J*) can serve as a biassing cause in the construction of adaptive niches, high entropy (*H*) acts as a biassing cause by determining how effectively these niches are utilised. A high *H* can function as a form of 'cognitive biassing,' where the organism's cognitive flexibility allows it to more effectively exploit the niches that its syntropy has enabled it to construct. This adaptability could influence the direction of both cognitive and biological evolution. For example, an organism with high syntropy might be more likely to construct complex tools or social systems, while an organism with high entropy would be better at adapting to the use of these tools or systems. This dual capacity would make it more likely for these advantageous traits to be passed onto future generations.

VI. Cognitive Neuroscience & Network Theory

Brain Network Modularity

Computational neuroscience of brain function has converged on some key ideas which are outlined below.

Constraints and Optimisation

- Both neural and computational systems face constraints between topological complexity of network connectivity and physical wiring cost.
- Optimising modularity can maximise topological complexity while adhering to cost constraints of physical embedding.

Integration and Segregation

- There is evidence that the human brain uses modular network architectures balancing integration and segregation. This maximises the functional complexity in the brain while minimising connection costs.
- Modularity facilitates specialisation of function in densely connected modules with sparse inter-modular links which are crucial for global integration and communication.

Roles of Connector Hubs

- Connector hubs link modules and promote global efficiency.
- Provincial hubs are crucial within modules for specialised, localised processing.

- Both connector and provincial hubs are proposed to play key roles in maintaining an optimal modular architecture by tuning connectivity within and between communities.
- ‘Rich Clubs’ are a set of highly interconnected hubs that facilitate efficient communication across the entire network. In the context of the metastability Θ function , key nodes of the Salience Network as well as other prefrontal areas act as a rich club, controlling boundary regions and coordinating multiple modules and networks.

Brain Network Modularity & Small World Topologies

Small-World Properties

- Modular networks can exhibit small-world properties if they have some inter-modular connections that provide shortcuts between modules.

Role of Connector Hubs in Small-World Networks

- The presence of connector hubs, which link different modules, enables modular networks to achieve small-world organisation.

Functional Specialisation and Global Integration

- Modularity provides the underlying architecture for functionally specialised processing. But the presence of connector hubs and a small-world topology allows globally integrated processing and information flow across the modular structure.

Brain Network Criticality & Self-Organised Criticality (SOC)

Computational neuroscience research also supports the following claims concerning brain network criticality:

Near Criticality States

- Functional brain networks operate near a critical point between order and randomness.
- Theoretically, this fact has been related to the theory of self-organised criticality: The SOC construct originates from the field of statistical physics and was first introduced by Per Bak, Chao Tang, and Kurt Wiesenfeld in 1987. The idea was to explain how complex systems can spontaneously evolve to a critical point, poised at the edge of

chaos and order. At this critical state, the system exhibits a high degree of complexity, and small changes can lead to cascading effects with consequences of all sizes—often described by a power-law distribution.

Complex Dynamics and Information Processing

- Operating at criticality allows complex dynamics like scale free avalanches and long-range correlations that support optimal information processing.
- Scale-Free Dynamics: Criticality is often associated with power-law distributions and scale-free dynamics, allowing for a wide range of dynamic behaviours.

Role in Modular Networks

- Modular networks tuned to an optimal balance between segregation and integration may underpin how the brain achieves a critical state.
- Near-critical dynamics occur in modular networks with an intermediate level of inter-modular connections, and this small-world architecture can generate critical dynamics across both short and long time scales.
- At the critical point, the system is optimised for maximal information transfer and computational capability.

Resilience & Resource Efficiency

- Systems at criticality are robust to perturbations, meaning they can adapt to a wide range of conditions without drastically altering their behaviour.
- Critical systems are often more energy-efficient, balancing the cost of neural computation and communication.

Meta-stability

Definition

- Meta-stability is a dynamic computational regime where the brain's network states are neither fully stable nor chaotic. This transient nature allows for a rich exploration of state-space, enabling the system to adapt to a wide range of tasks and environmental conditions.

Relation to criticality

- While both meta-stability and criticality involve the brain operating near a point of optimal computational capability, they emphasise different aspects of this balance. Criticality focuses on scale-free dynamics and optimal information transfer, whereas meta-stability highlights the system's ability to rapidly reconfigure its network states for task-specific demands.

Role in Modularity

- In modular networks, meta-stability emerges from an optimal balance between segregation and integration. It provides a computational mechanism for coordinating activity across spatial and temporal scales, allowing for both local specialisation and global integration.

Functional Implications

- Meta-stability facilitates quick and flexible transitions between different functional states. This is crucial for tasks requiring multi-modal integration or rapid adaptation to new information. It also provides a basis for balancing the trade-off between exploration and exploitation in decision-making and learning.

Neural Network Architectures & General Intelligence (*g*)

Modular Organisation and Intelligence

- Individual differences in intelligence are associated with differences in the modular organisation of functional brain networks.
- Specific mental abilities emerge from modular network communities that enable specialised, segregated information processing.

Small-World Topology and Broad Abilities

- Broad abilities like crystallised and fluid intelligence depend on the overall small-world topology and modularity tuning that balances segregation and integration.
- The capacity to transition between network states underlies general intelligence and relies on both provincial and connector hubs.

Dynamic Reconfiguration and General Intelligence

- Dynamic reconfiguration of this modular small-world network is critical for general intelligence.
- Modularity tuning may be a key mechanism of general intelligence, allowing the brain to achieve and maintain near-critical dynamics that support efficient information processing and complex neural computations.

Specialized Regions and Their Roles

- In more intelligent individuals, the Anterior Insula (AI), part of the Salience Network, shows higher between-module connectivity and is optimised for integrating and propagating information across different modules.
- In contrast, the Medial Superior Frontal Gyrus (SFG) and Temporo-Parietal Junction (TPJ), linked to the default mode network (DMN), show higher within-module connectivity. These regions may help to reduce the influence of potentially interfering information on goal-directed processing.

Criticality & Metastability and Fluid Intelligence

- Research suggests that individuals with higher fluid intelligence scores exhibit neural dynamics that are closer to a state of criticality. The critical dynamics associated with higher fluid intelligence are mostly observed in the prefrontal cortex and inferior parietal cortex.
- There is an association between maximal synchronisation entropy and individual variability in scale-free avalanche activity in a state of near criticality. This is a measure of the complexity and flexibility of neural architecture, linked to fluid intelligence and working memory.
- High resting state metastability in cognitive control networks (frontoparietal and dorsal attention) plays an important role in general intelligence across all tasks, and it is increased during task performance.

Efficiency and Flexibility

- The efficiency with which an individual's brain can switch from a resting to a task-based configuration correlates with intelligence. This could be interpreted as a measure of the system's ability to balance stability and flexibility, which is central to the concept of metastability.

Network Control Theory (NCT)

NCT is a branch of control theory that extends classical control theory to complex networks. Control theory is a multidisciplinary area of engineering and mathematics that deals with the behaviour of dynamical systems and how their behaviour can be modified by the use of feedback. In the context of brain networks, Network Control Theory provides a framework for understanding how the structure of the network (i.e., how neurons or brain regions are connected to each other) can influence its dynamics (i.e., how signals or information flows through the network).

Key NCT Concepts Applied to Brain Networks

- **Average Controllability**: This measures the ease with which a node (or brain region) can steer the system into many easily reachable states. In the article, regions with high average controllability are often densely connected hubs.
- **Modal Controllability**: This measures how a node can move the system to difficult-to-reach states. In the context of the brain, these are often weakly connected areas that are crucial for complex cognitive tasks.
- **Boundary Controllability**: This refers to the ability of a node to facilitate the integration or segregation of different modules or communities within the network. In the brain, these are often areas that lie at the boundary between different functional communities.

Research applying Network Control Theory to the study of brain networks has shown:

- **Default Mode Network**: Brain areas with high *average controllability* are those that can move the brain to many easily reachable states. These regions are often densely connected hubs, particularly in the default mode network (DMN).
- **Frontoparietal Networks**: Brain areas with high *modal controllability* are important for switching the brain between functions that require significant cognitive effort. These are usually weakly connected and are often found in cognitive control systems like the frontoparietal (and cingulo-opercular) networks.
- **Salience & Attention Control Networks**: Brain areas with high *boundary controllability* are important for integrating or segregating information across different cognitive processes. These are particularly enriched in nodes of the Salience Network (SN) and dorsal and ventral attention control networks. Regions include the anterior cingulate, rostral middle frontal, lateral orbitofrontal, frontal pole, medial orbitofrontal, and superior frontal areas. The SN, with key nodes in the anterior insula

and anterior cingulate, has rich connections with these boundary control areas and could serve as a central boundary control hub. It plays a crucial role in detecting and orienting attention, is known for its role in switching between the DMN and fronto-parietal networks, and is well-positioned to integrate or segregate information from different networks, a hallmark of boundary control.

Whole Brain Functional Networks

We can hypothesise the following relationships between the FPN, DMN, salience network and different network architectures:

Frontoparietal Network (FPN)

- The FPN (and its subnetworks) is critical for fluid intelligence and adaptive problem solving. The FPN is thought to underlie the ability to transition between network states, enabling access to difficult-to-reach states that support cognitive flexibility.
- The FPN is among the most costly in terms of metabolic energy usage. This aligns with the proposed role of the FPN in cognitively demanding tasks requiring transitions between network states.
- The FPN contains a high proportion of connector hubs, which facilitate transitions between network states and cognitive flexibility.
- Connector hubs in the FPN could link to diverse brain modules to enable global information integration.
- Provincial hubs may also facilitate within-module computational processes during complex cognition.

Default Mode Network (DMN)

- The DMN is associated with crystallised intelligence and accessing knowledge/experience through easy-to-reach network states. The DMN contributes to mutual interactions between cognitive processes that support crystallised abilities.
- In association with sensori-motor and memory networks, the DMN accesses provincial hubs to facilitate within-module processing and memory consolidation & retrieval.
- A highly modular architecture optimises the DMN for memory and automation.
- Connector hubs may link DMN modules to the FPN and sensori-motor networks when coordination is needed.

Salience Network (SN)

- The SN is specialised in detecting and filtering salient stimuli, thus playing a pivotal role in determining what information is relevant and should be attended to. Through its rich club of densely interconnected hubs, the SN can rapidly relay signals between the FPN and DMN, effectively coordinating these networks based on the salience processing.
- When integrated with prefrontal areas, the SN may form the neural substrate for a metastability function, dynamically adapt the balance between the FPN, DMN, memory and sensori-motor networks, depending on the context. For example, it could prioritise the FPN during problem-solving tasks and the DMN during introspective or memory-based tasks.
- The SN is central in the rich club of densely interconnected hubs for global communication and connector hubs in this rich club can rapidly relay signals between other control networks.

References

The Bayesian Brain

Papers that discuss the brain as a Bayesian inference machine. Starting with the Bayesian framework sets the stage for understanding the probabilistic nature of brain function. This is central to understanding the Free Energy Principle (FEP), its extensions and the FEP+.

- Doya, K. (2007). Bayesian brain: Probabilistic approaches to neural coding. MIT Press.
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Conceptual foundations. Trends in Cognitive Sciences, 14(7), 287-292. https://doi.org/10.1016/j.tics.2010.05.004
- Hohwy, J. (2013). The predictive mind. Oxford University Press.
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. Trends in Neurosciences, 27(12), 712-719. https://doi.org/10.1016/j.tins.2004.10.007
- Seth, A. K. (2015). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In P. Ford Dominey & V. Mueller (Eds.), Robots that talk and listen (pp. 89-114). De Gruyter.
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. Trends in Cognitive Sciences, 17(11), 565-573. https://doi.org/10.1016/j.tics.2013.09.007
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. Science, 331(6022), 1279-1285. https://doi.org/10.1126/science.1192788

Information Theory

Papers that discuss the mathematical and information-theoretic underpinnings of the FEP and FEP+. These references help clarify the mathematical foundations of the FEP+.

- Cover, T. M., & Thomas, J. A. (2006). Elements of information theory (2nd ed.). Wiley.
- Millidge, B., Tschantz, A., & Buckley, C. L. (2021). A mathematical walkthrough and discussion of the free energy principle. Neural Computation, 33(11), 2814-2855. https://doi.org/10.1162/neco_a_01433
- Shannon, C. E. (1948). A mathematical theory of communication. The Bell System Technical Journal, 27(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Thermodynamics

These references introduce the physical basis for understanding the FEP+.

- Brillouin, L. (1953). The negentropy principle of information. Journal of Applied Physics, 24(9), 1152-1163. https://doi.org/10.1063/1.1721463
- Collell, G., & Fauquet, J. (2018). Brain activity and cognition: A connection from thermodynamics and information theory. Journal of Cognitive Science, 19(4), 365-391. https://doi.org/10.17791/jcs.2018.19.4.365
- England, J. L. (2013). Statistical physics of self-replication. Journal of Chemical Physics, 139(12), 121923. https://doi.org/10.1063/1.4818538
- Kondepudi, D., & Prigogine, I. (2015). Modern thermodynamics: From heat engines to dissipative structures. John Wiley & Sons.
- Schrödinger, E. (1944). What is life? Cambridge University Press.
- Wolpert, D. H. (2019). The stochastic thermodynamics of computation. Journal of Physics A: Mathematical and Theoretical, 52(19), 193001. https://doi.org/10.1088/1751-8121/ab0850

Brain Energetics

Papers focusing on the energy landscapes in neural networks and how they relate to probability distributions. This serves as a bridge between the thermodynamic principles and their specific application in neural systems.

- Breakspear, M. (2017). Dynamic models of large-scale brain activity. Nature Neuroscience, 20(3), 340-352. https://doi.org/10.1038/nn.4497
- Gu, S., Cieslak, M., Baird, B., et al. (2018). The energy landscape of neurophysiological activity implicit in brain network structure. Scientific Reports, 8(1), 2507. https://doi.org/10.1038/s41598-018-20832-1
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, 79(8), 2554–2558. https://doi.org/10.1073/pnas.79.8.2554
- Sengupta, B., Stemmler, M. B., & Friston, K. J. (2013). Information and efficiency in the nervous system—a synthesis. PLoS Computational Biology, 9(7), e1003157. https://doi.org/10.1371/journal.pcbi.1003157

Life Sciences & Homeostasis

Introducing this biological 'embodied cognition' context early on helps to ground the FEP and FEP+ in real-world living systems.

- Cannon, W. B. (1932). The Wisdom of the Body. W. W. Norton & Company.
- Sterling, P., & Laughlin, S. (2015). Principles of neural design. The MIT Press.
- Varela, F. G., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organisation of living systems, its characterization and a model. Biosystems, 5, 187–196. https://doi.org/10.1016/0303-2647(74)90031-8

Extended Evolutionary Synthesis & Multiplier Effects :

This extends the biological context to contemporary evolutionary theory, providing a broader framework for understanding the FEP+.

- Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. Psychological Review, 108(2), 346-369. https://doi.org/10.1037/0033-295X.108.2.346
- Laland, K. N., et al. (2015). The extended evolutionary synthesis: Its structure, assumptions and predictions. Proceedings of the Royal Society B: Biological Sciences, 282(1813), 20151019. https://doi.org/10.1098/rspb.2015.1019
- Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., & Strassmann, J. E. (2015). The extended evolutionary synthesis: Its structure, assumptions and predictions. Proceedings of the Royal Society B: Biological Sciences, 282(1813), 20151019. https://doi.org/10.1098/rspb.2015.1019
- Laland, K. N., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. Evolutionary Ecology, 30(2), 191-202. https://doi.org/10.1007/s10682-015-9770-4
- Pigliucci, M., & Müller, G. B. (2010). Evolution: The extended synthesis. MIT Press.

Brain Functional Networks & Network Theory

Seminal papers that discuss how different brain networks (e.g. Frontoparietal, Default Mode and Salience/Prefrontal) are involved in free energy minimisation and the meta-functions of intelligence. This serves as a transition into brain-based interpretations of the FEP+.

- Alderson, T. H., Bokde, A. L. W., Kelso, J. A. S., Maguire, L., & Coyle, D. (2020). Metastable neural dynamics underlies cognitive performance across multiple behavioural paradigms. Human Brain Mapping, 41(12), 3212–3234. https://doi.org/10.1002/hbm.25009
- Alm, P. (2021). The dopamine system and automatization of movement sequences: A review with relevance for speech and stuttering. Frontiers in Human Neuroscience, 15. https://doi.org/10.3389/fnhum.2021.661880
- Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the 1/f noise. Physical Review Letters, 59(4), 381–384. https://doi.org/10.1103/PhysRevLett.59.381
- Barbey, A. K. (2018). Network neuroscience theory of human intelligence. Trends in Cognitive Sciences, 22(1), 8-20. https://doi.org/10.1016/j.tics.2017.10.001
- Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative cognition and brain network dynamics. Trends in Cognitive Sciences, 20(2), 87–95. https://doi.org/10.1016/j.tics.2015.10.004
- Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q., Fink, A., Qiu, J., Kwapil, T. R., Kane, M. J., & Silvia, P. J. (2018). Robust prediction of individual creative ability from brain functional connectivity. Proceedings of the National Academy of Sciences of the United States of America, 115(5), 1087–1092. https://doi.org/10.1073/pnas.1713532115
- Bertolero, M. A., Yeo, B. T. T., Bassett, D. S., & D'Esposito, M. (2018). A mechanistic model of connector hubs, modularity and cognition. Nature Human Behaviour, 2(11), 765-777. https://doi.org/10.1038/s41562-018-0420-6
- Bassett, D. S., Greenfield, D. L., Meyer-Lindenberg, A., Weinberger, D. R., Moore, S. W., & Bullmore, E. T. (2010). Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. PLoS Computational Biology, 6(4), e1000748. https://doi.org/10.1371/journal.pcbi.1000748
- Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition. In The Cognitive Neurosciences (4th ed., pp. 45-56). MIT Press.
- Deco, G., Kringelbach, M. L., Jirsa, V. K., & Ritter, P. (2011). The dynamics of resting fluctuations in the brain: Metastability and its dynamical cortical core. Journal of Nonlinear Science, 21(5), 423-463. https://doi.org/10.1007/s00332-010-9130-3
- Dixon, M. L., De La Vega, A., Mills, C., Andrews-Hanna, J. R., Spreng, R. N., Cole, M. W., & Christoff, K. (2018). Heterogeneity within the frontoparietal control network and its relationship to the default and dorsal attention networks. Proceedings of the National Academy of Sciences of the United States of America, 115(7), E1598-E1607. https://doi.org/10.1073/pnas.1715766115

- Gu, S., Pasqualetti, F., Cieslak, M., Telesford, Q. K., Yu, A. B., Kahn, A. E., Medaglia, J. D., Vettel, J. M., Miller, M. B., & Bassett, D. S. (2015). Controllability of structural brain networks. Nature Communications, 6, 8414. https://doi.org/10.1038/ncomms9414
- Hilger, K., Ekman, M., Fiebach, C. J., & Basten, U. (2017). Intelligence is associated with the modular structure of intrinsic brain networks. Scientific Reports, 7(1), 16088. https://doi.org/10.1038/s41598-017-15795-7
- Khundrakpam, B. S., Reid, A., Brauer, J., Carbonell, F., Lewis, J., Ameis, S., Karama, S., Lee, J., Chen, Z., Das, S., Evans, A. C., & Brain Development Cooperative Group (2013). Developmental changes in organisation of structural brain networks. Cerebral Cortex, 23(9), 2072–2085. https://doi.org/10.1093/cercor/bhs187
- Menon, V. (2011). Large-scale brain networks and psychopathology. In The Wiley-Blackwell Handbook of Mood Disorders (pp. 1-20). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119975568.ch1
- O'Callaghan, C., Walpola, I. C., & Shine, J. M. (2021). Neuromodulation of the mind-wandering brain state: The interaction between neuromodulatory tone, sharp wave-ripples and spontaneous thought. Philosophical Transactions of the Royal Society B: Biological Sciences, 376(1817), 20190699. https://doi.org/10.1098/rstb.2019.0699
- Okujeni, S., & Egert, U. (2023). Structural modularity tunes mesoscale criticality in biological neuronal networks. The Journal of Neuroscience, 43(14), 2515–2526. https://doi.org/10.1523/JNEUROSCI.1420-22.2023
- Power, J. D., et al. (2011). Functional network organization of the human brain. Neuron, 72(4), 665-678. https://doi.org/10.1016/j.neuron.2011.09.006
- Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., Moodie, C. A., & Poldrack, R. A. (2016). The dynamics of functional brain networks: Integrated network states during cognitive task performance. Neuron, 92(2), 544–554. https://doi.org/10.1016/j.neuron.2016.09.018
- Snyder, W., Uddin, L. Q., & Nomi, J. S. (2021). Dynamic functional connectivity profile of the salience network across the life span. Human Brain Mapping, 42(14), 4740–4749. https://doi.org/10.1002/hbm.25581
- van den Heuvel, M. P., & Sporn, O. (2011). Rich-club organisation of the human connectome. The Journal of Neuroscience, 31(44), 15775-15786. https://doi.org/10.1523/JNEUROSCI.3539-11.2011
- Walter, N., & Hinterberger, T. (2013). Self-organized criticality as a framework for consciousness: A review study. Frontiers in Human Neuroscience, 7, 1-15. https://doi.org/10.3389/fnhum.2013.00420

- Zabelina, D. L., Colzato, L., Beeman, M., & Hommel, B. (2016). Dopamine and the Creative Mind: Individual Differences in Creativity Are Predicted by Interactions between Dopamine Genes DAT and COMT. PloS one, 11(1), e0146768. https://doi.org/10.1371/journal.pone.0146768
- Xu, L., Feng, J., & Yu, L. (2022). Avalanche criticality in individuals, fluid intelligence, and working memory. Human brain mapping, 43(8), 2534–2553. https://doi.org/10.1002/hbm.25802

Cognitive Intelligence & Emotional Intelligence

Seminal works on general intelligence, cognitive ability and emotional intelligence, which is the focus of the FEP+.

- Barbey, A. K. (2018). Network neuroscience theory of human intelligence. Trends in Cognitive Sciences, 22(1), 8-20. doi: 10.1016/j.tics.2017.11.007
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54(1), 1-22. https://doi.org/10.1037/h0046743
- Gallese, V., Eagle, M. N., & Migone, P. (2007). Intentional attunement: mirror neurons and the neural underpinnings of interpersonal relations. Journal of the American Psychoanalytic Association, 55(1), 131–176. doi: 10.1177/00030651070550010601.
- Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. Behavioral and Brain Sciences, 30(2), 135-154. doi: 10.1017/S0140525X07001185
- Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The ability model of emotional intelligence. doi: 10.1037/amp0000103
- McGrew, K. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. Intelligence, 37(1), 1-10. doi: 10.1016/j.intell.2008.08.004
- Spearman, C. (1904). General intelligence objectively determined and measured. American Journal of Psychology, 15(2), 201-292.

Reinforcement, Motivation and Flow

Seminal publications on reinforcement learning and motivation, which help ground ‘flow’ and motivational aspects of the FEP+.

- Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. Harper & Row.
- Dietrich, A. (2004). Neurocognitive mechanisms underlying the experience of flow. Consciousness and Cognition, 13(4), 746-761. https://doi.org/10.1016/j.concog.2004.07.002
- Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In Flow and the foundations of positive psychology (pp. 239-263). Springer Netherlands.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary Educational Psychology, 25(1), 54-67. https://doi.org/10.1006/ceps.1999.1020
- Schultz, W. (2015). Neuronal reward and decision signals. In The Oxford Handbook of Computational Neuroscience (pp. 1-29). Oxford University Press. doi: 10.1093/oxfordhb/9780195393798.013.001

Intersubjectivity & Extended Cognition

Works that consider the boundaries of cognitive systems as 'extended' into external tools and technologies, a key concept in the FEP+.

- Clark, A., & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7-19. https://doi.org/10.1093/analys/58.1.7
- Clark, A. (2008). Supersizing the mind: Embodiment, action, and cognitive extension. Oxford University Press.
- Gallese, V., Eagle, M. N., & Migone, P. (2007). Intentional attunement: mirror neurons and the neural underpinnings of interpersonal relations. Journal of the American Psychoanalytic Association, 55(1), 131–176. https://doi.org/10.1177/00030651070550010601
- Kirchhoff, M. D., & Kiverstein, J. (2019). Extended consciousness and predictive processing. Philosophical Explorations, 22(2), 123-141. https://doi.org/10.1080/13869795.2019.1570479

Subjectivism vs Objectivism & Realism vs Constructivism Ontologies

Papers that discuss the ontological assumptions and implications of the FEP+. These references encourage the reader to think critically about the broader implications and assumptions underlying cognition.

- Clark, A., & Chalmers, D. (1998). The extended mind. Analysis, 58(1), 7-19. https://doi.org/10.1093/analys/58.1.7
- Clark, A. (2008). Supersizing the mind: Embodiment, action, and cognitive extension. Oxford University Press.
- Kirchhoff, M. D., & Kiverstein, J. (2019). Extended consciousness and predictive processing. Philosophical Psychology, 32(2), 161-184. https://doi.org/10.1080/09515089.2018.1553889

The Free Energy Principle and Its Extensions

Foundational papers that introduce and elaborate on the Free Energy Principle itself. With the foundational concepts in place, the reader is now prepared to review the FEP and its extensions leading to the FEP+.

- Badcock, P. B., Friston, K. J., & Ramstead, M. J. (2019). The hierarchically mechanistic mind: A free-energy formulation of the human psyche. Physics of Life Reviews, 31, 104-121. https://doi.org/10.1016/j.plrev.2019.01.005
- Badcock, P. B., Ramstead, M. J. D., Sheikhbahaee, Z., & Constant, A. (2022). Applying the Free Energy Principle to Complex Adaptive Systems. Entropy, 24(1), 1-19. https://doi.org/10.3390/e24010001
- Friston, K. J. (2010). The free-energy principle: A unified brain theory? Nature Reviews Neuroscience, 11(2), 127–138. https://doi.org/10.1038/nrn2787
- Friston, K. J. (2013). Life as we know it. Journal of the Royal Society Interface, 10(86), 20130475. https://doi.org/10.1098/rsif.2013.0475
- Friston, K. J. (2019). A free energy principle for a particular physics. Journal of Chemical Physics, 150(4), 040901. https://doi.org/10.1063/1.5058277
- Friston, K. J., & Kiebel, S. J. (2009). Predictive coding under the free-energy principle. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1521), 1211-1221. https://doi.org/10.1098/rstb.2008.0300
- Hohwy, J. (2021). Self-supervision, normativity and the free energy principle. Synthese, 199(1), 29-53. https://doi.org/10.1007/s11229-020-02622-2
- Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. Journal of The Royal Society Interface, 15(138), 20170792. https://doi.org/10.1098/rsif.2017.0792
- Ramstead, M. J., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. Physics of Life Reviews, 24, 1-16. https://doi.org/10.1016/j.plrev.2017.12.005

- Parr, T., Da Costa, L., & Friston, K. (2019). Markov blankets, information geometry and stochastic thermodynamics. Entropy, 21(10), 1003. https://doi.org/10.3390/e21101003
- Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. Progress in Neurobiology, 134, 17-35. https://doi.org/10.1016/j.pneurobio.2015.09.001

Philosophy of the Free Energy Principle

Works that tackle philosophical questions and criticisms related to the FEP. These references address some of the deeper, more conceptual questions and criticisms that arise around the FEP.

- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204. https://doi.org/10.1017/S0140525X12000477
- Clark, A. (2015). Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press.
- Sánchez-Cañizares, J. (2021). The Free Energy Principle: Good Science and Questionable Philosophy in a Grand Unifying Theory. Entropy, 23(2), 238. https://doi.org/10.3390/e23020238
- Klein, C. (2018). What do predictive coders want? Synthese, 195(6), 2597-2621. https://doi.org/10.1007/s11229-017-1547-9
- Ramstead, M. J. D., Kirchhoff, M. D., & Friston, K. J. (2019). A tale of two densities: Active inference is enactive inference. Adaptive Behavior, 27(1), 3-18. https://doi.org/10.1177/1059712318818515
- Kirchhoff, M. D., & Kiverstein, J. (2019). Extended consciousness and predictive processing. Philosophical Psychology, 32(2), 161-184. https://doi.org/10.1080/09515089.2018.1553889

Meta-Functions of Evolved Intelligent Systems

Papers that discuss higher-order functions and properties of intelligent systems according to the FEP+, such as entropy minimisation, syntropy maximisation and metastability. The references introduce higher-order concepts that build on the foundational and extended principles.

- Baars, B. J. (1993). A cognitive theory of consciousness. Cambridge University Press.

- Barbey, A. K. (2018). Network neuroscience theory of human intelligence. Trends in Cognitive Sciences, 22(1), 8-20. https://doi.org/10.1016/j.tics.2017.10.001
- Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence. The Behavioral and Brain Sciences, 30(2), 135-154. https://doi.org/10.1017/S0140525X07001185
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40, e253. https://doi.org/10.1017/S0140525X16001837
- Negro, P. J. (2019). Negative Entropy and Inference in Living Organisms. In A Blueprint for the Hard Problem of Consciousness (pp. 30-35). Bentham Science Publishers. https://doi.org/10.2174/9781681087785119010006
- Schneider, W., & Pimm-Smith, M. (1997). Consciousness as a message aware control mechanism to modulate cognitive processing. In J. Cohen & J. Schooler (Eds.), Scientific Approaches to Consciousness (pp. 271-287). Lawrence Erlbaum Associates.
- Schrödinger, E. (1944). What is life? The physical aspect of the living cell. Cambridge University Press.