Distributed Science

The Scientific Process as Multi-Scale Active Inference

Authors

Francesco Balzan [1,2*] (francesco.balzan3@unibo.it)
John Campbell [3]
Karl Friston [4,5]
Maxwell J. D. Ramstead [4,5]
Daniel Friedman[6,7]
Axel Constant[5,8]

Affiliation

1. Department of Computer Science, Alma Mater Studiorum, University of Bologna, Italy
2. Department of Computer Science, University of Pisa, Italy
3. Independent Researcher
4. Wellcome Centre for Human Neuroimaging, University College London, London WC1N 3AR, UK
5. VERSES AI Research Lab, Los Angeles, California, 90016, USA
6. Active Inference Institute
7. Department of Entomology and Nematology, University of California, Davis, USA.
8. Department of Engineering and Design, School of Engineering and Informatics, The University of Sussex, Brighton, UK

*Corresponding author

Keywords

Free-energy principle, active inference, scientific practice, Artificial Intelligence, meta-science, collective intelligence, cultural evolution, distributed cognition

Abstract

The scientific process plays out in a multi-scale system comprising subsystems, each with their own properties and dynamics. For the practice of science to generate useful world models—and lead to the development of enabling technologies—practicing scientists, their theories, methods, dissemination, and infrastructure (e.g., funding and laboratories) must all fit together in an orchestrated manner. Scientific practice has broad societal implications that go beyond mere scientific progress: we base our decisions on theoretical (i.e., models and forecasts) and technological (e.g., vaccines and smartphones) scientific advances. This paper applies the free energy principle to provide a multi-scale description of science understood as evidence-seeking processes in a nested hierarchy of living (biological and behavioural) and epistemic (linguistic) structures. This allows us to naturalise the scientific process—as distributed self-evidencing—in terms of dynamics that can be read as inference or Bayesian belief updating; i.e., processes that maximize the evidence for a generative model of the sensed and measured world. The ensuing meta-theoretical approach dispels the notion of science as truth-pointing and foregrounds inference to the best explanation—as evinced by the beliefs of scientists and their encultured niche. Crucially, it furnishes a way of simulating the practice of science, which may have a foundational role in the next generation of augmented intelligence systems. Epistemologically, it also addresses some key questions; e.g., is science a special? And in what ways is scientific pursuit an existential imperative for all beings? These questions may be foundational in how we use and design intelligent systems.

1. Introduction

This paper argues for a distributed view of science understood as the activity whereby knowledge is produced by human agents whose coordinated action forms an ecosystem of intelligence, read as Bayesian belief updating. The view on offer integrates what we call the “modern” conception of science as well as the “non modern” conception to provide an account of science both grounded on traditional theories of knowledge and on anthropological theories of knowledge.

1.1 The moderns

The origins of "modern science" can be traced back to the 17th century when Francis Bacon was looking for a distinction between the objective knowledge achieved via the scientific method and the subjective knowledge derived from the metaphysics of mediaeval philosophers. He highlighted the role of induction in enabling humans to eliminate subjective priors when constructing empirical knowledge. To achieve this, both Bacon [1] and later John Stuart Mill [2] argued that scientists must examine the world impartially, meaning they must neutrally (with no prior hypothesis or theory in mind) observe regularities of the world, until a universal statement can be induced. Induction alone, however, could not ensure objectivity. At best, it concealed science’s human factor under a veil of idealised neutrality. The modern view of science—that emerged during the 17th century—thus fostered an idealised version of the works of scientists as an unbiased activity to generate objective theories that enable us to get a grip on reality.

Induction alone fails to recognize the influence of biases and previous hypotheses in the observation process, as lately argued by [3] and [4]. Additionally, it does not allow for the attainment of universal truths, as it is based on the assumption that if something is true in several observed instances (the sun has always risen), it will be true in all instances (the sun will always rise) (see Hume's concept of inductive fallacy [5]). Finally, it does not preclude the development of rival scientific theories based on identical evidence, as the evidence available to scientists is always consistent with multiple theoretical frameworks [6]. This characterises the science of the “moderns” [7], which has inherited a fallacious dichotomy between nature and society, leading to a misunderstanding of the ways in which scientific knowledge is shaped by human and non-human (e.g., material and technological) agencies.

1.2 The non-moderns

The "nature-culture divide" is the idea that there is a clear distinction between the natural world and human society. According to Latour, the traditional separation between nature and society has led to an oversimplification of both domains, with nature being seen as a static, passive entity and society as the active force shaping it. An alternative approach to defining science is to view it as a collaboration between human and non-human agencies [8]. In this view, the world is knowable to the extent that we have tools, methods, and theories that scaffold scientific knowledge by imposing top-down constraints on scientific activity [9]. This historical perspective, referred to as "non-modern science" (following [7]) for the purposes of this article, can be traced back to the philosophical productions of William Whewell and William Stanley Jevons, with their introduction of the hypothetico-deductive method of science ([10];[11]) and to Charles Sanders Peirce, the first to propose abductive inference, combining aspects of deduction and induction [12]. Abduction involves a two-stage process in which one generates sets of hypotheses and then infers, based on data and specific constraints (e.g., simplicity, coherence), which proposed hypothesis is most likely. In Peirce's account, abduction introduces new hypotheses into the scientific process, deduction determines the logical implications derivable from these hypotheses, and induction subjects these implications to testing by evidence in order to achieve a scientific generalisation. Lorenzo Magnani extended Peirce’s concept of abduction by proposing the notion of “manipulative abduction” [13]; referring to a form of reasoning where hypotheses are generated and tested through interactions with the physical world, often through the manipulation of physical objects or systems. This concept challenges the modern view of scientific reasoning as a purely mental or symbolic process, and emphasises the role of material practices and embodied cognition in scientific discovery.

1.3 This paper

The non-modern picture of scientific knowledge is rich and appears to capture most of the key reasoning steps involved in the production of scientific knowledge. Abduction allows for the logical deduction of implications from a predetermined subset of hypotheses that have been carefully selected out of a process of hypothesis formation that involves an interaction between a variety of agents, humans and material alike. In Bayesian epistemology, this process is sometimes referred to as the context of discovery and the context of justification [14]. Salmon [14] proposes to read the context of discovery as that which sets prior probabilities based on social and psychological facts for the context of justification; wherein scientists employ conditionalization and other methods of Bayesian model comparison to test their hypothesis. For Salmon, under the Bayesian interpretation, prior expectations (probabilities) are picked up by scientists from their interaction with the world (experiments), which inevitably affects not only how they evaluate different hypotheses in light of new evidence (justification) but also the active, creative process of hypothesis formation through abduction. The context of discovery, or the history of science, as Salmon would put it, is central to scientific thinking. A key aspect of the Bayesian interpretation of the relationship between evidence and hypotheses is the notion that prior experiences, which inform prior beliefs, influence how an epistemic agent views an observation in relation to hypotheses.

The Bayesian interpretation reconciles the modern and non-modern views of science by accounting for the way the socio-material and historical context feeds into scientific practice understood as a manipulative abductive process to shape knowledge production. Bayesian accounts provide a good formal story of how social and psychological factors come to shape the production of scientific knowledge, or context of justification. But how do scientists themselves shape their socio-material and historical context? How do institutions change through scientific activity? What are the top-down and bottom-up relations between the way scientists initialise and learn their priors and how their work comes to shape the institutions from which subsequent generations of scientists acquire their priors? We believe that to get the full Bayesian picture of how science is progressed by individuals and how it is shaped by scientific communities—which themselves are shaped by the collective operations of their constituents—the Bayesian approach to scientific knowledge needs to detail the belief-based (cognitive) mechanisms whereby the context of discovery and the context of justification interact.

In line with Bayesian approaches to animal cognition [15–17], the theory of active inference—which is a Bayesian theory of cognition—proposes a formal definition of intelligence as an activity of generating evidence for beliefs about the structure of the world [18]. Under this definition, intelligence is a process that integrates the sampling of evidence (e.g., action when applied to animal systems), the inference over hidden states causing evidence (e.g., perception when applied to animal systems), and the update of priors and likelihood (e.g., learning when applied to an animal system). These Bayesian belief updating processes take place at "nested" spatial and temporal scales along the hierarchy of self-organising systems (i.e., from individual cells to human communities). As systems scale up (e.g., as individuals form communities), the range of intelligent behaviour (i.e., the "cognitive light cones" [19]) also increases, which increases cognitive sophistication.

Distributed science, under active inference, describes networks of intelligent agents whose "cognition" across scales can be described in terms of approximate Bayesian inference. The term cognition here refers to the three basic processes described by active inference: action, perception, and learning. Scientific cognition, in terms, refers to how cognition is leveraged to produce scientific knowledge. We focus on two general scales of scientific cognition: (i) the individual scale, at which scientific cognition operates through individual-level cognitive functions (e.g., how executive functions allow an agent to seek, acquire and produce knowledge) (discussed in section 2) and (ii) the collective scale, at which scientific cognition operates through institutional processes of scientific communities (e.g., how extended and embodied cognitive operation of scientists come to generate communities that embody scientific knowledge) (discussed in section 3).

The two scales of scientific cognition can be viewed as integrated hierarchically—the collective level supplying top-down control (i.e., empirical priors or inductive biases) on the individual level, and the individual level providing bottom-up drivers of collective scientific cognition. The hope is that our account of science could provide a quantitative, mechanistic framework for studying collective intelligence, and for understanding how individual cognitive processes can give rise to science in complex socio-technical systems.

2. Individual scientific cognition

The pursuit of comprehending, and possibly improving cognition and intelligence through computational simulations can be traced back to Aristotle, whose syllogism demonstrates his intuition of intelligence as a form of symbol manipulation and computation. Following this research path, in many instances, scientific progress has been signed by the development of tools for computational simulations as they provide a conceptual framework to facilitate the exploration of processes and a methodology for conducting experiments with process-based theories [20]. As a fruitful example, cognitive neuroscience has seen tremendous advances in the last decades thanks to the explorative deployment of minimally complex and maximally accurate computational models of cognitive processes ([21]). In this section, we present Thagard's attempt to (reflexively) apply the same methodology to understand a particular type of cognition: scientific cognition, opening the path to the Cognitive Science of Science field. In 2.1, we highlight the limitations of his models as a starting point for our proposal: equipping Bayesian epistemology (section 2.2) with some active inference moves (section 2.3) to more optimally model the human ability to generate scientific knowledge under specific top-down, socio-cultural constraints (section 3).

2.1. Computational Models of Scientific Cognition

In his paper "Scientific Cognition: Hot or Cold?" ([22]), Paul Thagard explores individual scientific cognition within the framework of computational philosophy of science, integrating the perspectives of history and philosophy of science, artificial intelligence, and cognitive sciences ([23]). By recognising the potential of cognitive science methodologies for testing meta-scientific hypotheses related to scientific cognition, Thagard adopts the connectionist framework to simulate scientific "cold" rational reasoning and to test his meta-scientific hypothesis of "explanatory coherence" [24]. Thagard's model, ECHO (Explanatory Coherence), simulates the establishment of various scientific theories (e.g., oxygen combustion theory and phlogiston, the Darwinian theory of natural selection and creationism) by incorporating seven principles that establish local coherence relations among hypotheses and other propositions, including: coherence through explanation, being explained, participating in explanations of other propositions, and offering analogous explanations. The model treats hypothesis evaluation as a constraint satisfaction problem, implementing the principles through a connectionist program. It creates a network of units representing propositions based on inputs about explanatory relations, while coherence and incoherence are encoded using excitatory and inhibitory links (see figure 1). ECHO offers an algorithm that integrates theory evaluation based on explanatory breadth, simplicity, and analogy, representing hypothesis plausibility through node activation levels.

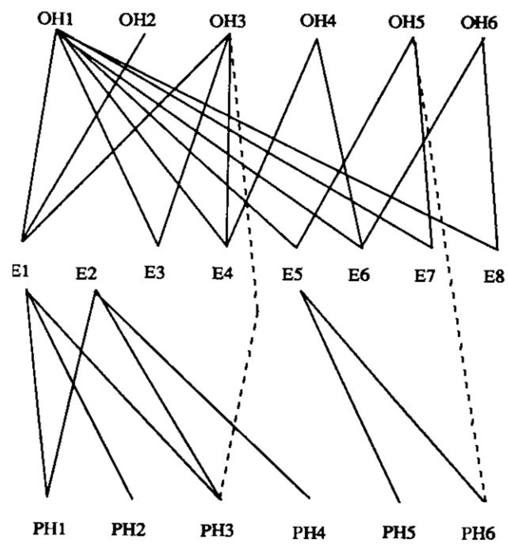

Figure 1. The ECHO model network representing Lavoisier's argument (1862). E1-E8 are evidence units. OH1-OH6 are units representing hypotheses of the oxygen theory; PH1-PH6 represent the phlogiston hypotheses. Solid lines are excitatory links; dotted lines are inhibitory. From [25].

ECHO has also been applied by Ranney [26] in the educational field, comparing students' belief updating based on evidence in physics with ECHO's performance, resulting in similar outcomes and rendering ECHO a reasonable model of individual, evidence-based reasoning. However, it is important to acknowledge that the ECHO connectionist model does not account for the psychological aspects of scientific practice. To address this limitation, Thagard and colleagues designed HOTCO and Motiv-PI ([27]), winch incorporate what Thagard defined the "hot" aspects of scientific cognition (i.e., the emotional and psychological variables that influence scientific practice alongside rationality) encompassing what has been defined as "motivated inference":

> In Motiv-PI, the system biases inference to favour generalizations that are positively relevant to a representation of the self. For example, the generalization "Extraverts are successful" is relevant to you if one of your motivations is to be successful [22].

Thagard's distinction between "cold" and "hot" scientific cognition resonates with our differentiation between modern and non-modern views of science. Thagard keeps these two aspects separate to compare their performance with historical data of scientific evolution [22]. He concludes that historical evidence suggests a rational model (ECHO) as the most appropriate, indicating a predominance of "cold" cognition in scientific thinking.

Notably, the author claims that this predominance is partly due to social factors since even scientists driven by personal motivations of success and fame must present research to the scientific community in terms of experimental and theoretical merits [22]. By stating this, the author seems to suggest that to fully understand the functioning of science and the dynamics of scientific cognition, it is necessary to consider the dual interaction between the individual and the collective level. In this paper, we take that provocation seriously.

Therefore, aligned with the non-modern view, we claim that Thagard's models have two main limitations: (i) they fail to capture scientific cognition as a distributed cognitive process involving a complex network of human and non-human actors, and, while they contemplate the influence of individual preferences, (ii) they ignore the history of science or the context of discovery (defined by the collective dynamics of scientific communities). To overcome these limitations, we suggest employing active inference as a means to model distributed scientific cognition at multiple scales. The claim here is that active inference enables us to make significant progress in the ambitious goal set forth by Thagard and his colleagues; namely, to find the artificial neural correlates (a computational representation) of scientific cognition. This research has the potential to enhance our understanding of both human scientific cognition and the capabilities of artificial intelligence systems. In our introduction to the active inference framework, we first explore Bayesian epistemology as a comprehensive interpretative lens for understanding scientific cognition. This approach allows us to reconcile both the modern ("cold") and non-modern ("hot") perspectives on science, providing a unified framework for analysis amenable to be equipped with active inference.

2.2. Bayesian Epistemology

Bayesian epistemology provides a formal probabilistic framework that enables the reconciliation of prior beliefs with new data, allowing for the evaluation of evidence. This approach is particularly valuable in scientific hypothesis testing, as it incorporates the prior probability of a hypothesis being true, a factor that has traditionally been viewed as possessing a non-rational element. This prior probability essentially represents a "best guess" or an initial assumption, which has prompted philosophers of science to exercise caution when applying Bayesian methods to scientific practice. A prominent scholar in the field, Wesley Salmon, addressed this concern by proposing that the historical development of science itself provides valuable insights into the rationality of Bayesian reasoning within scientific contexts. According to Salmon, the accumulation of empirical evidence over time, coupled with the systematic testing and revision of hypotheses, serves to refine and update the prior probabilities [14]. The history of science becomes an essential element of scientific functioning as it describes how prior probabilities, based on social and psychological factors, directly influence not only the context of discovery (i.e., finding new scientific hypotheses) but also the context of justification where scientists employ methods such as conditionalization and Bayesian model comparison to test their hypotheses. We now briefly rehearse a simple example to show how history sets prior probabilities influencing individual scientific cognition.

The heliocentric model of the solar system, which places the Sun at the centre of the motion of our local planetary system rather than the Earth, initially faced significant opposition due to the prevailing cultural and religious beliefs of the time. Let's consider two hypotheses: H1 is the Earth-centred model, and H2 is the Sun-centred model. The evidence, E, is the observational data, such as the apparent retrograde motion of planets. Before the evidence is considered, due to strong cultural and religious beliefs in an Earth-centred universe, the prior probabilities might look like this:

- P(H1) = High
- P(H2) = Low

When the evidence (E) comes in, we consider the likelihood of that evidence given each hypothesis. That is, how likely would we see this evidence if the hypothesis were true?

- P(E | H1) = Low (as the Earth-centred model struggled to account for observations without complex additions like epicycles)
- P(E | H2) = High (as the Sun-centred model accurately predicted the motion of the planets)

We then update the probabilities based on this new evidence using Bayes' theorem. For the Sun-centred model, it would look like this:

$$P(H2 \mid E) = P(E \mid H2) * P(H2) / P(E)$$

Despite the initial low prior for H2, the strong evidence in favour of it (high P(E | H2)) would result in an updated probability (posterior) that is higher than the initial one. As more evidence accumulates, our beliefs update, sometimes overcoming even strong initial biases. However, it's important to note that if the prior belief in H1 is extremely strong, it may take a significant amount of evidence and attention to meaningfully shift the belief towards H2. In a real-world context, this can reflect factors like societal resistance to paradigm shifts in understanding.

Responding to the first limitation of Thagard’s computational model; in the Bayesian framework, the material and technological context play a role in providing new evidence, refining old evidence, and sometimes even creating entirely new areas of inquiry. In fact, technologies through which we gather evidence primarily impact the likelihoods – that is, the probability of the evidence given the hypothesis, P(E | H). Let's continue with the heliocentric vs. geocentric model example. Before the invention of the telescope, the evidence available was limited and sometimes even misleading. For instance, the naked-eye observation that the Sun and stars have periodic visibility in Earth’s sky was (and is) consistent with a geocentric model. With the invention of the telescope, humans could gather more accurate and detailed observations. This new tool provided evidence such as the phases of Venus and the moons of Jupiter, which were highly inconsistent with the geocentric model but well-explained by the heliocentric model. In Bayesian terms, this would drastically decrease P(E | H1) and increase P(E | H2), thereby shifting the posterior probabilities in favour of the heliocentric model.

The Bayesian approach enables us to bridge the gap between modern and non-modern views of science by incorporating the influence of technologies and prior probabilities from the external environment. This resolution of apparent contradictions highlights the reciprocal influences between these perspectives. Theoretical models play a crucial role in guiding the constructive activities of scientists, as they strive to make sense of the world. Conversely, the outcomes of these activities, such as engineering or experimentation, influence the theoretical models of other scientists by shaping the prior probabilities and likelihoods of hypotheses in light of new evidence. To further facilitate this merging process between modern and non-modern views of science and embrace the multi-scale and dynamic nature of scientific cognition, we propose integrating a standard Bayesian interpretation of scientific knowledge construction with the cognitive-based, scale-free approach provided by the Free Energy Principle (FEP) and the active inference framework.

2.3. Individual Scientific Cognition as Active Inference

The Free Energy Principle (FEP) is built on the elementary assumption that living systems are characterised by resisting entropic decay: that is, they do not dissipate as do many transient phenomena (e.g., tornadoes). Living systems persist and thrive by frequenting a limited set of states with low entropy compared to all possible states [28]: e.g. a living system programmed to live in water, will avoid the surprising (given its original program) eventuality of finding itself on land. From a statistical point of view, this means that living systems manifest self-organising, non-equilibrium steady-state dynamics that we can associate with the phenotype of a living system [29]. The FEP claims that systems can be read as leveraging an internal probabilistic generative model. Generative models furnish a probabilistic mapping between external states of the world and internal states of the system [30]. Under the FEP, generative models are statistical models entail by particular systems for the production of adaptive behaviours via the selection of specific policies. The FEP tells us that external states are conditionally independent of internal states and that this conditional independence rests on the maintenance of a Markov Blanket between the internal and external states of the system. Markov Blankets, therefore, allow one to define a particular system of interest (e.g., particle or person) and to characterise its exchange with the environment via active and sensory states (constituents of a Markov Blanket); where active states influence, but are not influenced by external states and sensory states influence but are not influenced by internal states [28]; [31]; [32]. This characterization of self organisation leads to an elementary description of living systems and information-processing entities with biological processes as implemented computations [33].

In the FEP-theoretic approach, the adaptive behaviour of systems is implemented as active inference. In active inference, the particle or agent selects the course of actions that it believes will evince its characteristic (i.e., preferred) sensory states. The generative model can be cast as encoding probabilistic mapping from causes in the external environment (i.e., external states) to the sensory states that they generate. The generative model can therefore be decomposed into a likelihood (the probability of sensory consequences, given external causes) and a prior (the probability of external causes). Equivalently, the generative model can be decomposed into a posterior (the probability of external causes, given sensory consequences) and the marginal likelihood of those sensory consequences. This marginal likelihood is also known as model evidence, where the negative logarithm of model evidence constitutes self information (in information theory) or, more simply, surprise (a.k.a., surprisal). In other words, surprise scores the implausibility of a particular sensory outcome given the agent or generative model in question. Inference, in this setting, corresponds to maximising marginal likelihood or minimising surprise. Crucially, the expected surprise or self information is the entropy of sensory consequences that is implicitly minimised when minimising surprise. This brings us back to the proclivity of particular (usually biotic) systems to resist increases in entropy. In short, to persist in characteristic states is to infer the causes of sensory consequences. In the FEP, surprise is associated with an upper bound called variational free energy. This is a useful quantity because it is a functional of an agent's sensory data and some (Bayesian) beliefs about the causes of those data encoded by the systems internal states. As an upper bound, variational free energy is always greater than surprisal, which means that minimising variational free energy implicitly minimises surprise; to the extent the bound is a good approximation. This leads to the notion of approximate Bayesian inference that provides a tractable (Bayesian) mechanics for belief updating.

Active inference—premised on a generative model—entails the selection of actions that generate expected sensory data, which counts as evidence for the existence of the system; namely, self-evidencing [34]. The minimisation of surprisal or free energy operates over different timescales, which correspond to the unobserved (external) causes of sensory states that are hidden behind an agent's Markov blanket:

1. State estimation: beliefs about the latent or hidden states that generate sensory outcomes are optimised via perceptual inference;
2. Parameter learning: model parameters, which encode contingencies and statistical regularities, are optimised via learning;
3. Structure learning: the structure of the generative model itself can be optimised via model selection [35].

The three types of hidden causes are optimised at distinct temporal scales and each scale both inherits from—and contextualises—the scale below. Generally speaking, Bayesian belief updates are closely related: Bayesian model selection (structure learning) determines which parameters are relevant to the task at hand; in online learning, the specific value of these model parameters is learned iteratively (parameter learning); which, in turn, optimises state estimation via perceptual inference [35]. Under the FEP, beliefs about each of these kinds of unknowns are optimised (i.e., the posterior distributions are estimated) through a process ascribable to approximate Bayesian inference which, under active inference, implies a dual methodology:

1. Perception and learning: updating model parameters via maximising model evidence (i.e., minimising variational free energy). In other words, generating knowledge from evidence.
2. Active sampling and selection: sampling sensory data to minimise surprise and expected surprise (i.e., uncertainty). In other words, generating evidence from knowledge.

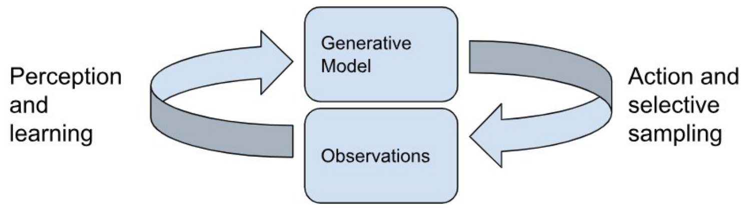

Figure 2. The dynamic underpinning systems' self-organization via active inference over a probabilistic generative model of the external world: the generative model underwrites the production evidence for its own plausibility, via action and selective sampling; the resulting observations are used to update the parameters and the structure of the generative model via perception (state estimation) and (parameter or structure) learning.

Technically, beliefs about hidden causes are updated by minimising variational free energy. Similarly, beliefs about action are updated by minimising expected free energy; namely, the free energy expected under beliefs about the (observable) consequences of action. Interestingly, minimising expected free energy can be expressed as complying with the principles of optimum experimental design and Bayesian decision theory. This follows because expected free energy is the expected information gain plus expected value, where value is the expected marginal likelihood.

We characterise individual scientific cognition as the implementation of such a dual methodology for approximate Bayesian inference. In particular, we identify the context of discovery, or history of science, in which hypotheses' prior probabilities forming scientists' generative models of reality are picked up in their interaction with the external world, with the process of generating knowledge from evidence (perception and learning); and the context of justification, wherein such hypotheses are tested, with the process of generating evidence from knowledge (action and selective sampling). Inevitably, the context of discovery, or history of science, affects the justification process (i.e., how individual scientists evaluate hypotheses in light of new evidence) needed for the update of new knowledge in the generative model which, in turn, affects the context of discovery of new scientific hypotheses (abduction). Figure 3 shows the result of translating the same dual methodology for individual scientific cognition.

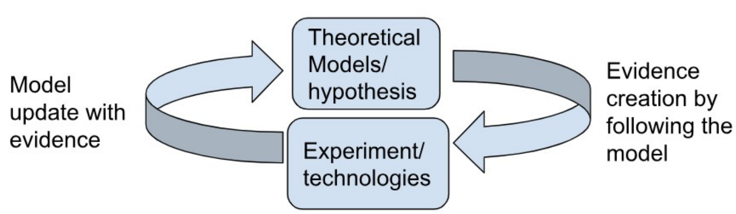

Figure 3. Individual scientific cognition as a dialectic process:

1) Model update via evidence (left arrow): when new data or evidence is collected, we update our models or hypotheses to better fit that evidence. This is the essence of Bayes' theorem: given new evidence, how

should we adjust our beliefs (probabilities) about our hypotheses? This process of model update is a central tenet of the scientific method, where theories must adapt in the face of new empirical findings.

2) Evidence creation by following the model (right arrow): on the other hand, our models or hypotheses also guide the collection of new data. The model suggests what evidence would be relevant and what kind of observations should be made to test the model. In this way, our current beliefs (models) shape the direction of scientific inquiry. Technological advancements often play a key role in this process, as they expand our capabilities to gather new types of data.

Importantly, the modern view of science encapsulates scientists' endeavour to construct models mirroring or aligning with a designated system [36], such as the heliocentric model of the solar system. This view concentrates on the context of justification and—akin to Thagard's "cold" cognitive models—confines scientific cognition description to the creation and updating of models through perception and learning (left arrow), without addressing its dynamic and distributed constituents. The subjoined illustration delves into the interplay between the observation-generating process (the generative process) and the theoretical model embodied by the scientist to (passively) infer the underlying causes of these observations, effectively recapitulating the generative process's causal structure (the generative model). This graphical depiction encapsulates the unidirectional process of constructing scientific models, as proposed by the modern view of science. Within the framework of active inference, this corresponds to the initial phase of the dual methodology that systems employ to optimize various aspects of the generative model: the extraction of knowledge from evidence through perception and learning.

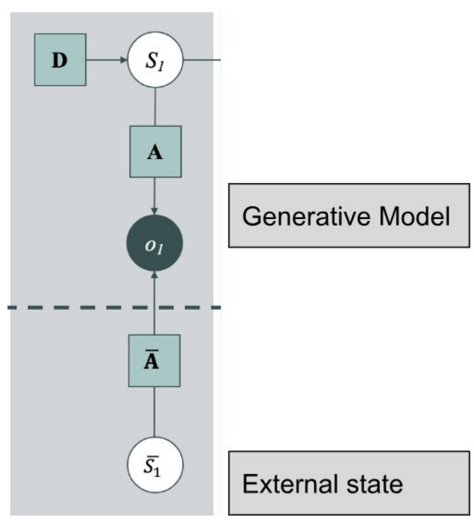

Figure 4. Modern science. Graphical representation of the generative model/external state (GM/ES) interaction during state inference and parameter fine tuning (i.e., perception and learning). The upper part of the image represents the generative model embodied by the scientist in its interaction with the world. D is the prior probability distribution weighting a set of mutually exclusive and exhaustive hypotheses forming a model, whose probabilities sum to 1. The probabilities forming this distribution are calculated from prior observations. S_i is the most probable scientific theory or hypothesis: the most probable state out of the true posterior. A is the posterior probability: the probability of observing o1 given Si ($P(o_1|S_i)$). This calculation is made for all S_i forming the model and the probabilities are normalised by dividing each by the marginal likelihood over the full model to form the likelihood ratio. The updated posterior probability

is then P_i x A_i. (o_1) is the scientific observation generated by the dynamics of the external states, described in the lower part. The ES is composed of a hidden state (S_1 bar) which causes the observation detected by the system via sensory states (o_1). It is the “true”, “ideal” posterior distribution, the true state of affairs of the world that generated the scientific observables (p(s)). $\bar{A}$ is a parameter that maps states in the external world (ES) with observations.

However, as demonstrated earlier, within the active inference framework, the process of model updating (perception and learning) constitutes just half of the narrative. Active inference agents not only update their internal generative models but also actively orchestrate the production of evidence to substantiate these models (self-evidencing through action and selective sampling). For an individual [scientist] engaged in active inference, this translates to the deliberate selection of experimental configurations capable of supplying evidence that effectively diminishes uncertainty about competing hypotheses entertained under their generative model. Consequently, the non-modern perspective of science acknowledges scientists' manipulative endeavours as an integral facet of scientific cognition; designated in Bayesian epistemology as the context of justification.

In the realm of active inference, this involves constructing an expressive generative model with temporal depth, wherein forthcoming states of the world are inferred as hidden states, commencing from present and past observations. This is an important move because the generative model now encompasses the future. And the future depends upon action. In this setting, action now becomes a cause of observable consequences and, perhaps counterintuitively, has to be inferred. In turn, this leads to the notion of planning as inference in a general setting or, the issue of experimental design (and data selection) in a scientific setting.

Parr and Friston [31] introduced the notion of generalized free energy in which the expected free energy—that underwrites future action or experimentation—is combined with the variational free energy to furnish a single objective function that ensures the minimisation of surprise and expected surprise. As noted above, expected surprise has two aspects. The first reads expected surprise as uncertainty leading to actions that maximise information gain. This aspect is often couched in terms of information seeking and epistemic affordance. The second entails avoiding surprising outcomes with a small marginal likelihood. This aspect is often couched in terms of goal seeking and instrumental affordance. In terms of experimental design, this simply means that to be Bayes optimal—in the sense of active inference—is to solicit experimental data or observations that resolves the most uncertainty about (i.e., disambiguate) scientific hypotheses, while avoiding outcomes that would be uncharacteristic of the scientist in question (e.g., blowing herself up) or characteristically unscientific (e.g., unethical). See Box 1 for a formal summary of variational and expected free energy minimisation.

Box 1: active inference

Recent trends in theoretical neurobiology, machine learning and artificial intelligence converge on a single imperative that explains both sense-making and decision-making in self-organising systems, from cells [37] to cultures [38]. This imperative is to maximise the evidence (a.k.a., marginal likelihood) for generative (a.k.a., world) models of how observations are caused. This imperative can be expressed as minimising an evidence bound called variational free energy [39] that comprises complexity and accuracy [40]:

Free energy = model complexity – model accuracy

Accuracy corresponds to goodness of fit, while complexity scores the divergence between prior beliefs (before seeing outcomes) and posterior beliefs (afterwards). In short, complexity scores the information gain or cost of changing one's mind. This means Bayesian belief updating is about finding an accurate explanation that is minimally complex (c.f., Occam's principle). In an enactive setting—apt for explaining decision-making—beliefs about 'which plan to commit to' are based on the free energy expected under a plausible plan. This implicit planning as inference can be expressed as minimising expected free energy [41]:

Expected free energy = risk (expected complexity) – precision (expected accuracy)

Risk is the divergence between probabilistic predictions about outcomes, given a plan, relative to prior preferences. Precision is the expected accuracy (e.g., avoiding ambiguity such as noisy or dark rooms). An alternative decomposition is especially interesting from the perspective of the scientific process:

Expected free energy = expected cost – expected information gain

The expected information gain underwrites the principles of optimal Bayesian design [42], while expected cost underwrites Bayesian decision theory [43]. However, there is a twist that distinguishes active inference from expected utility theory. In active inference, there is no single, privileged outcome that furnishes a cost function. Rather, costs are replaced by uncharacteristic or surprising outcomes, quantified by their (log) marginal likelihood. In short, active inference appeals to two kinds of Bayes optimality and subsumes information and preference-seeking behaviour under a single objective function that scores epistemic and pragmatic affordances, respectively.

In Figure 5, we combine the modern and non-modern viewpoints of science into a cohesive probabilistic generative model illustrating the decision-making process of a scientist. Only through the interplay of both realms (discovery and justification) can the agent effectively execute "good science," thereby exhibiting adaptive behaviour and engendering adaptive outcomes (e.g., scientific theories and technologies).

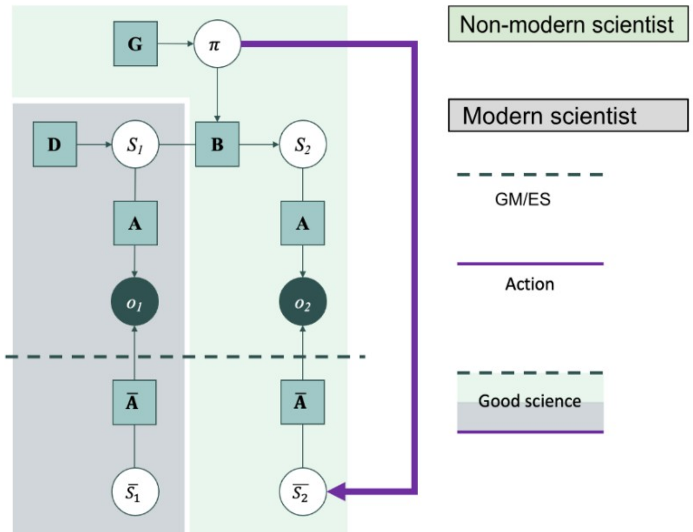

Figure 5. A simple generative model entailed by an active inference scientist, in which modern and non-modern views are integrated. The non-modern view (green) equips the modern view with temporal depth; enabling the scientist to plan her experiments prior to execution. Here G is the expected free energy functional and updates the posterior probability of various plans or policies π, which is used to select actions that solicit new data. π implies the selection of the right experiment, behaviour or available technology to produce evidence that maximises knowledge gain or disambiguates competing hypotheses (purple arrow). It is a quantity that we want to infer (e.g., which experiment should I run? Where should I look next?). Beliefs on how the world will change in response to actions are modelled via the transition matrix B (what will I observe, if I do that? What will I see if I look over there?).

By adding the future to the generative model, we implement planning as inference. This involves the evaluation of the expected free energy of alternative plans that include the expected information gain (e.g. how my scientific beliefs will change when faced with a scientific observation generated by my experiment?). Future states are defined in terms of probabilistic “beliefs” about the evolution of states based on the policy in question. Therefore, choosing a policy means choosing a state transition matrix that brings about states that resolve uncertainty. Where do policies come from and how do they influence scientific decisions? By answering these questions, in the next section, we provide evidence for the hypothesis proposed by Thagard:

> Because of an institutional commitment of science to experimental evidence and explanatory argument, science as a whole is able to transcend the personal goals of its fully human practitioners who acquire the motivation to do good experiments and defend them by rational argument. [22]

In other words, how does the institution of science contextualise scientists’ “hot” cognitive tendencies and guarantee the prevalence of “cold” cognition in the evolution of science?

3. Collective scientific cognition

We have seen above that the non-modern view of science—fuelled by the postmodern movement and pursued by authors such as [8] and [9]—recognizes science as a cultural phenomenon and as such, accepts it as a self-motivated, autodidactic endeavour. However, as intuited by Thagard, the socio-cultural aspect of science is also what “guarantees” the rationality of its cognitive components. In this section, we describe science as a social practice aimed at actively structuring the external world to make it measurable and knowable. Importantly, the activity of scientists is aimed at bringing about observations that match their expectations where expectations, in turn, are picked up from the external cultural niche over development and learning. In simpler words: education through our scaffolded socio-cultural niches constrains the possible behaviours of scientists and shapes what scientists expect the world to reveal [37] and in turn, scientists and engineers structure the external niche in ways that support the expected and preferred outcomes to come about.

3.1. Distributed Scientific Cognition and Niche Construction under the FEP

The distributed cognition paradigm has proven quite fruitful for comprehending various forms of sophisticated, human collective behaviours—from large ships' navigation [38] to the design of a new language [39]. With roots in Vygotsky’s *Mind in Society* [40], the distributed cognition paradigm suggests that human cognition goes beyond the boundaries of individuals to include the interaction with media in the external environment (e.g., other people and technologies). Such media are not passive tools for learning, as suggested by the constructivists, but instead are central and active components of cognition, as proposed by the non-modern approach to science [41] and, more generally, by proponents of extended mind and cognition [42,43].

In line with the distributed cognition paradigm, Latour suggested the concepts of "hybrid" and "network" as an alternative to the dichotomy between nature and society. He claims that our understanding of the world should be based on a mixture of human and non-human entities, which he calls "actants". These actants are part of networks or assemblages, through which they interact and shape each other. Giere [44] extends this analysis by noting that science represents an explicit case in which the boundaries of individual cognition are extended to a larger sociocultural system that, in turn, endows its individual sub-components with sophisticated computational capabilities, unimaginable for isolated agents. We have seen above that this renders the inferential cognitive process involved to be best understood as distributed between the scientist and the scientific niche constituted by other scientists, tools, affordances and deontic cues. In this section, we claim that the Bayesian Mechanics and the FEP are useful for formalising the distributed cognition approach to the practice of science, while still rendering justice to the modern and non-modern interpretations of science described above.

The FEP naturally offers scale-free heuristics to make sense (by simulating them) of complex, self-organising dynamics at different scales. The FEP has been applied to generate insight into structured, collective behaviours (e.g.,[32]; [45]; [46]; [47]) as forms of coordinated and distributed inference ([48]; [29]). Key to the scalability of FEP to collective behaviours is the idea that the areas of concern of agents—that is, the domain of their observations—can grow via higher-order pattern formation. The main process that has been posited to enable the extension of an area of concern is community formation [49]. In fact, experiments show that multicellular ensembles are able to extend their areas of concern by orders of magnitude in time and space, compared to isolated cells, by forming communities that share information (via various kinds of neural or bioelectric signalling; [50]), which lies outside the bounds of each agent's Markov blanket—thereby forming a kind of higher-order Markov blanket [49]. Social groups can attune to regularities that their constituent members would not plausibly pick on over one lifetime ([51]; [45]; [52]). For example, elderly elephants have memory capacities that enable them to guide their herd to distant sources of water in times of severe drought. More relevant to our interests here, we describe the process of scientific investigation as a process of cumulative knowledge and technology construction, which can be seen as extending the area of concern of human agents to a nearly unlimited scope—extending the boundaries of the human area of concern to the boundary of the observable universe.

As described above, with science, the world becomes more knowable through tools, methods, models, and technologies that are collectively constructed and offloaded into the external environment. We might resume this scaffolding process as *Scientific Niche Construction* (SNC) [53], in which, such a modified niche radically influences the way scientists expect the world to be and, consequently, drives their actions in the world which are aimed at generating the evidence to confirm their expectations or hypotheses [54]. As glimpsed above, these two processes are deeply intertwined in the practice of science: models guide the collection of new evidence, and that new evidence in turn refines the models. This ongoing cycle of hypothesis generation, evidence collection, and hypothesis revision is a key driver of scientific progress. Crucially for our argument, FEP-theoretic modelling can be leveraged to explain how the components of a distributed and composite system—like the one underwriting scientific cognition—are able to coordinate and form robust new patterns—and indeed, new emergent systems —at a superordinate scale ([55], [56]; [57], [46]. Such first-principles agent-based modelling approaches have the potential to describe existing scientific informational resources and active entities [58].

3.2. The Socio-Technical System of Science

FEP-theoretic models are (almost always[1]) formulated, explicitly or implicitly, in a multi-scale manner, and (usually) rest upon the formal tools that underwrite the study of adiabatic processes and the renormalisation group. Most formulations of the FEP appeal to timescale separation in order to define the states of things at all. The idea is rather simple. It is an empirical fact that nature manifests a nested, multiscale organisation: with small, fast things (e.g., atoms and molecules) coalescing into progressively larger, and slower things (e.g., crystals, biofilms, and organisms)—and so on, recursively and iteratively. One critical thing to

1 Note that so-called quantum formulations of the FEP (which are called quantum because they leverage the holographic principle and quantum information geometry—not because they apply to atomic scales) eschew the specification of a spacetime background; and therefore, they are scale-free, or scale-friendly, but not inherently multi-scale.

note about this nesting is that, as one ascends the nested scales of things, from the small and fast, to the large and slow, events in some sense “average out”, such that we can treat fast stable dynamics at one scale as random fluctuations at the next, superordinate scale. For example, the lifecycle of one individual blood cell happens so quickly—relative to the lifecycle of the organism in which it lives—that the particularities of its lifecycle can be considered as a random fluctuation relative to that of the organism. Similarly, the life cycle of a single scientific hypothesis within a scientist’s brain might have little epistemic relevance when referring to the evolution of science as a whole. Thus, the very states that make up a system are, in the FEP-theoretic context, defined implicitly in terms of the (spatial and temporal) scales at which it is meaningful to speak of a thing as a cohesive locus of states. In other words, things that exist physically, as separable things, can only really be said to exist to the extent that they change slowly enough—and with sufficiently stable and rich structure—to be reliably re-identified by an observer using the right reference frame (here the Free Energy Principle). In turn, slowly changing states are effectively treated as parameters of the generative model: by varying at a slower timescale, they in effect parameterise or modulate the flow of states (see section 3.1. for a more detailed explanation).

In Bayesian mechanics, the main mechanism that is proposed to enable the formation and maintenance of community formation (and pattern formation at superordinate scales more generically) is communication, premised on a shared generative model. The idea is that ensembles of agents that share the same—or similar enough—beliefs about the typical sensory consequences of an action will be able to figure out which role they play in a larger pattern. In particular, recent simulation work has suggested that the key to the emergence of stable higher-order patterns and structure is the endowment of agents with specific beliefs about group membership at the superordinate scale (e.g., [32]; [45]; [46]; [47] [29]). Importantly, to actually achieve such a higher-order formation, membership beliefs must be satisfied by evidence.

We claim here that science is such a higher-level ensemble of Markov blanketed systems. We claim, in particular, that science can be described as an emergent, partially independent system whose behaviour both constrains and is constrained by the behaviour of its constituent parts. Indeed, one can view the process of scientific investigation as an evolutionary process, leading to the selection of specific forms of existence and to the definition of particular constraints, or policies, which underpin the correct production of scientific knowledge and technologies. In that way, the socio-technical system of science and its infrastructure provide context (i.e., parametrise) to the inferences of the scientists that engage in scientific investigation, reinforcing the belief of being part of a higher system (See [46] regarding the spontaneous emergence of higher-level systems via the expectations of belonging). These “strange feedback loops”[2] that ensue from the circular coupling between agent and niche might, e.g., take the form of disruptive scientific technologies and theories that enhance individual cognitive adaptability and fitness. Every time someone uses some piece of technology or successfully enacts a policy or plan because of information that is embedded in the scientific context (either in other agents or

2 See [59] notion of the strange loop as a self-referential, hierarchical structure in which the levels are intertwined in such a way that the highest level leads back to the lowest level, creating a closed loop with no clear beginning or end.

in the environment), this serves as evidence for an implicit generative model that is entailed by higher-order information gathering. This is made possible because we share roughly the same set of cultural prior beliefs.

What is important for the goal of this section, is that it is exactly this feedback loop from the higher to the lower level that permits us to refer to science as a higher-order socio-technical system. In the words of Kirchhof and colleagues:

> The conservation of Markov blankets (of Markov blankets) at every hierarchical scale enables the dynamics of the states at one scale to enslave the (states of) Markov blankets at the scale below, thereby ensuring that the organization as a whole is involved in the minimization of variational free energy. It is thus only when the properties of the collective dynamics feed back into the scale below, forming a free energy-minimizing system at the scale of the whole system, that it is possible to talk meaningfully of ensemble Markov blankets—blankets whose self-evidencing dynamics result in an overall self-sustaining organization [32].

Crucially, the multi-scale self-evidencing dynamic is true both for the individual agent expecting to be part of a higher system and for the scientific system as a whole which, like any other inferential system, is aimed at producing adaptive, existing entities—one that emerges from the situated collective enactments of the denizens of a given niche. By producing evidence of its own existence, we can state that the scientific process might be interpreted and modelled as a partially independent, socio-technical system that implements self-evidencing dynamics. One that, arguably, has rediscovered and streamlined the methods that are used by natural systems in their self-evidencing. Analogously to an ant colony, the scientific system may itself be amenable to being described as a “Bayesian superorganism” ([60]; [61]). There are things that colonies know that nestmates don't. And there are things that scientific groups/communities know that individual researchers do not.

By integrating collective with individual scientific cognition we can now overcome the limitations of former computational models of scientific cognition (ECHO model) and respond to a contradiction that might emerge from a Bayesian integration of the modern and non-modern views of science: the historicity and contextuality of scientific practice (non-modern science) versus the (momentary) universality of its empirical outcomes (modern science). Reproducibility, testing by evidence and other top-down “rules” and constraints that emerged during the evolution of the scientific system parametrizes the actions of its lower-level components by selecting specific policies that scientists can pick up from their niches as deontic cues. By so doing, human communities are able to overcome the individual limitations of their basic components and to (at least partially) silence their priors in light of a higher-order, intelligent goal. Is this enough for arguing that the outcomes of scientific practice are, even if momentarily, of a superior status compared to other knowledge production systems?

3.3. Anticipating Brains are not (always) “Crooked Scientists”

Our argument rests in some sense on our ability to draw an analogy between the activities of anticipating brains, as described by the FEP, and the hypothesis-testing abilities of scientists, which was first proposed in its contemporary form by Helmholtz ([62,63] and later developed in the context of computational modelling of perception by [64]. Now, it has been argued that this analogy is flawed [65]: according to this argument, there is a deep disanalogy between the preference-driven manner in which living systems infer the causes of their sensations (Thagard’s “hot” cognition) and the objective, scientific manner in which scientists do so (“cold” cognition). This is because anticipating brains bring about their preferred data distributions, and they are unlike honest scientists—they are “crooked scientists”. In this view, the idea of perception as a kind of scientific hypothesis testing process is flawed, because scientists must take the evidence as it comes and refrain from skewing it, e.g., via selective sampling, to support particular hypotheses; whereas this biased data collection is precisely what is mandated by active inference.

The arguments exposed in the paper nuance this view, on two counts. For one, the idea that brains must either be like good scientists or have preferences about the data that they generate is, in our view, a false dichotomy. While we agree that, under the FEP, anticipating brains indeed act in such a way as to generate their preferred sensory data, we do not think that this undermines the analogy to scientific hypothesis testing—provided that this analogy be extended to consider the practice of “non-modern science”. The “crooked scientist” that is described by Bruineberg and colleagues [65] is just a scientist who acts in the world to gather evidence for her hypothesis. The real difference is made by the top-down influence of the higher-level, socio-technical system of science that feeds back into the activity of individual scientists by constraining their behaviour in virtue of deontological principles selected during the evolution of science.

Our view is that, despite such top-down constraints, scientists and scientific research groups are not neutral parties. For better or for worse, scientists usually do perform experiments in order to generate the evidence that would best disambiguate among their favourite hypotheses. Indeed, scientific communities are cultural communities, and they have a vested interest in confirming the hypotheses to which they have committed their careers. Scientists and their research groups routinely (almost on a yearly or biyearly basis) compete for resources in the space of research (namely, things like grant money, attention from the public and other scientists, room for publication in journals, etc.). Science is an evolutionary process of model selection. This provides us with a new vantage point on Planck’s famous statement that “science progresses one funeral at a time.” Some of these pragmatic issues with scientific practice, which seem to detract from its epistemic value, are brought to light and clarified in the account presented here. In our view, these properties are not merely bugs—rather, they are features of science as a form of existence. As such, we ought to expect that these pragmatic factors will play an important role in science.

The quasi-religious belief in the efficacy of scientific investigation—what has been called scientism—might actually be selected for, when we cast science as a multiscale evolutionary process. Socio-cultural phenomena like modern religions are good examples of systems that have been able to trigger and reinforce population-level beliefs about the importance of belonging to a higher-order system. If religious and political systems have prompted the acquisition of this belonging belief with the goal of securing their domination and power, scientific investigation as a collective system of inference generates a form of life that provides evidence for the expectation that one is indeed part of a higher-order system. Interestingly it does so, like religious practice, by rewriting the history of a community (in terms of a march towards scientific progress, as in the positivist philosophy of science) and by invoking unseen, hidden agents (the hypothetical causes of data inferred by scientists).

4. Conclusion

This treatment identifies scientific methodology as a manifestation of the FEP following the same essential tradition as all natural entities in accumulating and applying knowledge for existential purposes. This tradition may be observed, for example, in biology where natural selection plays the role of “making better models of the world” and in developmental biology the role of “following the model accurately” as well as in neuroscience, where belief updating—under generative models—with sensory information informs policy based behaviour and *vice-versa.* But human science is obviously distinct in some aspects from these prior forms of knowledge accumulation and application. First, science has a knowledge repository distinct from either (epi) genetics or neuronal generative models. Scientific knowledge is communal and is stored in repositories such as libraries or computer memories. Second, the role of updating scientific models is abstract, following the mathematics of Bayesian inference. What is called the iron law of science is that evidence and only evidence counts in updating scientific knowledge: a principle similar to the constrained maximum entropy principle that the only constraint on entropy or ignorance is evidence (for a generative model). This abstract and mathematical relationship between model updating and evidence, without requiring new tests, results in a streamlined inferential process as compared, to say, natural selection where the same variant hypothesis (allele variation) may be tested out repeatedly in the real world. Still, science depends on evidence provided through experimentation, but experiments need not be random, as in natural selection, but may be designed, through processes such as Bayesian experimental design, to produce evidence that is maximally effective or has maximal expected information gain when updating the model.

Another unique aspect of human science is the indulgently long leash by which we are tied to our existential master. As opposed to natural selection, where every retained characteristic is retained only if it is neutral or beneficial to reproductive success, human science and engineering not only indulge in neutral but costly traits such as space travel but also traits threatening our survival such as nuclear weapons. Perhaps science now has sufficient freedom to even erase the substrate on which it depends; It remains to be seen if this latitude is a blessing or a curse. But this latitude is an aspect of science’s ‘area of concern’ which has expanded to include phenomena within the entire observable universe. Now that science, along with everything else of concern, is coming to be understood in terms of the FEP, we may expect knowledge accumulation to undergo an acceleration as general knowledge is made easier as boundaries between disciplines are removed; if you know a little about the FEP, you know a little about everything. Given this further streamlining of the scientific endeavour, we may expect its area of concern to accelerate in depth as well as breadth.

It has become clear in many instances that the reason Bayesian inference produces superior scientific theories is that the generative models central to existing natural entities also form and function as Bayesian processes [66]. With this "radical conceptual revolution", the arena of internal models, such as life's genetic model, transitions from scientific constructs residing in theorists' brains to mechanistic Bayesian processes residing in actual life forms. Crucially, this revolution is the result of a highly streamlined method for evaluating scientific hypotheses of the external reality by testing throughout *in silico* simulations. From a meta-level of analysis, therefore, the FEP constitutes a novel abstract generative model through which scientific hypotheses are generated (abduction), while the active inference framework translates these hypotheses into synthetic entities (evidence creation by following the model) which are then confronted with their real-world counterparts. The mismatch (surprise) that emerges from the comparison is then used to update the original theoretical model (model update with evidence) (Figure 6). Therefore, the cultural practice of science, exemplifying the processes of evolutionary development, is able to carve out a space within existence through a cyclical evolutionary process that employs theory to engineer experiments, technologies and other cultural structures having a high degree of fitness, in the sense that they are able to proliferate.

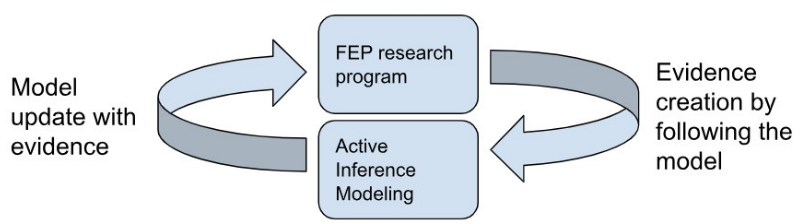

Figure 6. Translation of the dual methodology derived from the FEP to the FEP itself and its active inference implementation.

This approach consolidates the idea that existence is inference and provides a maximal expression of it, by arguing that physical existence and evolution themselves are continuous with this kind of hypothesis testing. In some metaphorical sense, the FEP can be taken as delivering on Carl Sagan's view that existence and evolution are "nature's way of coming to know itself." Across all levels of existing things, individual existence is best accomplished by faithfully following a particular generative model, one fine-tuned by its phylogenetic history, to produce evidence of that thing's existence. Processes of active inference ensure that hard-won knowledge for existence is applied as diligently as possible.

We have proposed to understand scientific practice in terms of collective inference premised on shared generative models, which consist largely of consensus hypotheses shared by the scientific community. This collective inference arguably results in the existence of at least two persistent entities. The first is the scientific community itself. Each generation of scientists inherits the model borne by their predecessors, and their contribution to its evidence accumulation and to the application of the bodies of knowledge that they entail is largely constrained to problem-solving at the model periphery, much as the accumulation of genetic knowledge is largely constrained to mutations solving adaptive problems at the model periphery. One of the main reasons that scientific investigation evolves is that it significantly—dramatically, even—extends the area of concern (cognitive light cone) of humans, all the way up to the limits of the observable universe and all the way down to the limits of observable thingness.

A second type of persistent entity produced by scientific generative models are cultural technologies, structures and activities. Engineering models, adapting scientific models for achieving specific outcomes, generate a large body of cultural entities. For example, engineering models adapt Turing's and Von Neumann's theoretical computational models to generate computers. Another aspect of the existential power of these scientific/engineering generative models is the existence of nearly 8 billion people on our planet, largely supported through the applications of these models.

By framing scientific inquiry within the context of nested active inference, a departure from 20th-century falsificationist epistemology towards a post-Popperian philosophy of science becomes evident. Falsificationism, propagated by Popper—which emphasized the progress of science and its corresponding demarcation criteria—can be considered a retrospective interpretation of genuine scientific practice. The shift has taken place from falsificationism to one focused on evidence-based model comparison. In this evolution, the validation of empirical theories is no longer contingent upon the potential falsification of a hypothesis through experimentation. Instead, the emphasis lies in comparing alternative explanations (models) that elucidate the generation of specific data, assessing the evidence each model garners from the analysed data.

Ultimately, this Bayesian and cognitive paradigm in scientific practice aligns with a pragmatic stance. The central objective here is the creation of models capable of guiding human endeavours in generating adaptive entities, such as technologies or policies. The mounting proliferation of successful scientific outcomes and entities, like artificially intelligent systems and technologies, could potentially serve as historical evidence supporting the hypothesis underpinning this perspective: that the evolution of science is integral to the universal enhancement of strategies for constructing and sustaining existence through evidence-based knowledge storage. Scientific investigation, as a natural process, is inherently subject to the same governing regularities that typify the physical world.

For a philosopher, this perspective resonates with a 'Darwinio-peripatetic-Platonic' conception of knowledge, wherein ideal forms, inaccessible in the empirical realm (the true distribution), are progressively approximated through empirical sampling of sensory outcomes and the process of learning via experimentation. Evolution bestows upon us a foundational framework or model structure (e.g., hierarchical levels) through the pruning of models over phylogeny, ready to be reconstructed on species-specific 'o'—thus, we seemingly 'rediscover' the ideal forms we are evolutionarily predisposed to rediscover over developmental stages.

5. Future research paths

We claim that an FEP-based computational model of scientific cognition could be used to predict (via simulation testing) which specific forms of interactions, affordances and top-down constraints (policies) might have led to, and further promote, the emergence of higher-level forms of intelligence. This might help us understand not only how science has emerged as an epistemic, evolutionarily stable strategy, but, importantly, how individuals come to infer knowledge content in the form of scientific statements about external states of affairs and thus clarify how scientists come to represent reality with models that experience minimal surprise when encountering data concerning that reality.

An FEP-based generative model of the socio-technical system of science can provide valuable insights into the dynamics and mechanisms of collective intelligence. We have seen that the model could help us understand how collective intelligence emerges from the interactions of heterogeneous actors throughout a shared (cognitive) niche. By examining the influence of hidden variables and external constraints on the system, we can identify factors that promote or hinder the emergence of effective collective scientific cognition. For instance, the model might reveal how certain funding structures or communication technologies enhance the community's collective problem-solving capabilities. By so doing, the model can provide insights into the structure and dynamics of the scientific community as a collectively intelligent system. This includes how information and influence flow through the community, how it responds to new evidence or challenges, and how it collectively updates its beliefs and strategies. These insights could inform the design of other collectively intelligent systems, such as crowdsourcing platforms or decentralized decision-making bodies.

Additionally, the integration of the evolutionary/systemic perspective with the developmental/psychological one might permit us to track the influence of a higher-level, socio-cultural system—like science—on the online cognitive functions of its elementary components (i.e., scientists) in terms of, for example, cognitive penetrability, rational inference and pro-social behaviours. Worth noting, the same approach can be used to assess and compare the impact of other sociocultural systems (e.g., a religious system) on the cognitive capabilities of its active sub-components.

Acknowledgement

The authors are grateful to VERSES for supporting open access publication of this paper. This work was supported by a ERC-2020-SyG, European Research Council Grant (XSCAPE, Agreement number 951631), Australian Laureate Fellowship project A Philosophy of Medicine for the 21st Century (Ref: FL170100160), by a Social Sciences and Humanities Research

Council doctoral fellowship (Ref: 752–2019-0065) and by PNRR M4C2 Investiment 1.3, Extended Partnership PE0000013 - "FAIR" - Spoke 8 "Pervasive AI" funded by the European Commission under the NextGen EU programme. KF is supported by funding for the Wellcome Centre for Human Neuroimaging (Ref: 205103/Z/16/Z) and a Canada-UK Artificial Intelligence Initiative (Ref: ES/T01279X/1).

Bibliography

1. Bacon F. The New Organon and Related Writings. Anderson FH, editor. New York: Liberal Arts Press; 1960. Available: https://philpapers.org/rec/BACTNO-2
2. Mill JS. A System of Logic. 1843. Available: https://en.wikipedia.org/w/index.php?title=A_System_of_Logic&oldid=1140577348
3. Popper KR. The Logic of Scientific Discovery, reprinted (1959). London: Hutchinson.
4. Kuhn TS. The structure of scientific revolutions. Chicago (university of Chicago press) 1962. 1962 [cited 1 Apr 2023]. Available: https://opus4.kobv.de/opus4-Fromm/frontdoor/index/index/docId/28136
5. Hume D. An Enquiry Concerning Human Understanding: A Critical Edition. Oxford University Press; 2000. Available: https://play.google.com/store/books/details?id=3Vp-0Y3Yz_cC
6. van Orman Quine W. Two Dogmas of Empiricism. In: Harding SG, editor. Can Theories be Refuted? Essays on the Duhem-Quine Thesis. Dordrecht: Springer Netherlands; 1976. pp. 41–64. doi:10.1007/978-94-010-1863-0_2
7. Latour B. We Have Never Been Modern. Harvard University Press; 1993. Available: https://play.google.com/store/books/details?id=xbnK8NzMsm4C
8. Latour B. Pandora’s hope : essays on the reality of science studies. Cambridge, Mass. [u.a.]: Harvard University Press; 2000.
9. Pickering A. The mangle of practice: time, agency, and science. Chicago: University of Chicago Press; 1995.
10. Whewell W. The Philosophy of the Inductive Sciences: Founded Upon Their History. J.W. Parker; 1840. Available: https://play.google.com/store/books/details?id=8xg2AQAAMAAJ
11. Stanley W. Jevons. The Principles of Science: A Treatise on Logic and Scientific Method. Macmillan and Co. 1874.
12. Peirce CS. The Essential Peirce: Selected Philosophical Writings. Indiana University Press; 1992. Available: https://play.google.com/store/books/details?id=T2weTOqdjqcC
13. Magnani L. Abduction, Reason and Science: Processes of Discovery and Explanation. Springer Science & Business Media; 2011. Available: https://play.google.com/store/books/details?id=UmlyBgAAQBAJ

14. Salmon W. Bayes's theorem and the history of science. Historical and Philosophical Perspectives of Science. 1970.

15. McNamara JM, Green RF, Olsson O. Bayes' theorem and its applications in animal behaviour. Oikos. 2006;112: 243–251. doi:10.1111/j.0030-1299.2006.14228.x

16. Richerson PJ. An integrated bayesian theory of phenotypic flexibility. Behav Processes. 2018. doi:10.1016/j.beproc.2018.02.002

17. Okasha S. The Evolution of Bayesian Updating. Philos Sci. 2013;80: 745–757. doi:10.1086/674058

18. Friston KJ, Ramstead MJD, Kiefer AB, Tschantz A, Buckley CL, Albarracin M, et al. Designing Ecosystems of Intelligence from First Principles. arXiv [cs.AI]. 2022. Available: http://arxiv.org/abs/2212.01354

19. Levin M. Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds. Front Syst Neurosci. 2022;16: 768201. doi:10.3389/fnsys.2022.768201

20. Gilbert N. A Simulation of the Structure of Academic Science. Sociol Res Online. 1997;2: 91–105. doi:10.5153/sro.85

21. Doerig A, Sommers RP, Seeliger K, Richards B, Ismael J, Lindsay GW, et al. The neuroconnectionist research programme. Nat Rev Neurosci. 2023. doi:10.1038/s41583-023-00705-w

22. Thagard P. Scientific Cognition: Hot or Cold? In: Fuller S, de Mey M, Shinn T, Woolgar S, editors. The Cognitive Turn: Sociological and Psychological Perspectives on Science. Dordrecht: Springer Netherlands; 1989. pp. 71–82. doi:10.1007/978-94-015-7825-7_4

23. Thagard P. Computational Philosophy of Science. MIT Press; 1988. Available: https://play.google.com/store/books/details?id=Mw-Yj95_BfgC

24. Thagard P. Extending explanatory coherence. Behav Brain Sci. 1989;12: 490–502. doi:10.1017/S0140525X00057319

25. Thagard P. Explanatory coherence. Behav Brain Sci. 1989;12: 435–467. doi:10.1017/S0140525X00057046

26. Ranney MA. Changing naive conceptions of motion. Diss Abstr Int. 1975;49. Available: https://psycnet.apa.org/fulltext/1989-53039-001.pdf

27. Thagard P, Kunda Z. Hot cognition: mechanisms of motivated inference. Proceedings of the Annual Meeting of the Cognitive Science Society. 1987. pp. 753–763.

28. Friston K. Life as we know it. J R Soc Interface. 2013;10: 20130475. doi:10.1098/rsif.2013.0475

29. Ramstead MJD, Hesp C, Tschantz A, Smith R, Constant A, Friston K. Neural and phenotypic representation under the free-energy principle. Neurosci Biobehav Rev. 2021;120: 109–122. doi:10.1016/j.neubiorev.2020.11.024

30. Friston K, Heins C, Ueltzhöffer K, Da Costa L, Parr T. Stochastic Chaos and Markov Blankets. Entropy. 2021;23. doi:10.3390/e23091220

31. Parr T, Friston KJ. Generalised free energy and active inference. Biol Cybern. 2019;113: 495–513. doi:10.1007/s00422-019-00805-w

32. Kirchhoff M, Parr T, Palacios E, Friston K, Kiverstein J. The Markov blankets of life: autonomy, active inference and the free energy principle. J R Soc Interface. 2018;15. doi:10.1098/rsif.2017.0792

33. Fields C, Levin M. Scale-Free Biology: Integrating Evolutionary and Developmental Thinking. Bioessays. 2020;42: e1900228. doi:10.1002/bies.201900228

34. Hohwy J. The self-evidencing brain. Noûs. 2016;50: 259–285.

35. Friston KJ, Da Costa L, Parr T. Some Interesting Observations on the Free Energy Principle. Entropy . 2021;23. doi:10.3390/e23081076

36. Godfrey-Smith P. Models and fictions in science. Philos Stud. 2009;143: 101–116. doi:10.1007/s11098-008-9313-2

37. Friston K, Levin M, Sengupta B, Pezzulo G. Knowing one's place: a free-energy approach to pattern regulation. J R Soc Interface. 2015;12. doi:10.1098/rsif.2014.1383

38. Veissière SPL, Constant A, Ramstead MJD, Friston KJ, Kirmayer LJ. Thinking through other minds: A variational approach to cognition and culture. Behav Brain Sci. 2019;43: e90. doi:10.1017/S0140525X19001213

39. Winn J, Bishop CM. Variational Message Passing. 2005 [cited 16 Oct 2023]. Available: https://www.jmlr.org/papers/volume6/winn05a/winn05a.pdf?q=variational

40. Ramstead MJD, Sakthivadivel DAR, Heins C, Koudahl M, Millidge B, Da Costa L, et al. On Bayesian mechanics: a physics of and by beliefs. Interface Focus. 2023;13: 20220029. doi:10.1098/rsfs.2022.0029

41. Friston KJ, Daunizeau J, Kilner J, Kiebel SJ. Action and behavior: a free-energy formulation. Biol Cybern. 2010;102: 227–260. doi:10.1007/s00422-010-0364-z

42. Lindley DV. On a Measure of the Information Provided by an Experiment. aoms. 1956;27: 986–1005. doi:10.1214/aoms/1177728069

43. Berger JO. Statistical Decision Theory and Bayesian Analysis. Springer Science & Business Media; 2013. Available: https://play.google.com/store/books/details?id=1CDaBwAAQBAJ

44. Kuhn TS. The structure of scientific revolutions. University of Chicago Press; 1962.

45. Hutchins E. Cognition in the Wild. MIT Press; 1995. Available: https://play.google.com/store/books/details?id=CGIaNc3F1MgC

46. Wundt WM. Probleme der Völkerpsychologie. Kröner; 1921.

47. Vygotsky LS. Mind in Society: The Development of Higher Psychological Processes. Harvard University Press; 1980. Available: https://play.google.com/store/books/details?id=Irq913lEZ1QC

48. Latour. Visualization and cognition. Knowledge and society. Available: http://hci.ucsd.edu/10/readings/Latour(1986).pdf

49. Clark A. Surfing Uncertainty. 2016. doi:10.1093/acprof:oso/9780190217013.001.0001

50. Clark A, Chalmers D. The Extended Mind. Analysis. 1998;58: 7–19. Available: http://www.jstor.org/stable/3328150

51. Giere RN. Scientific Cognition as Distributed Cognition. 2015 [cited 17 Apr 2023]. Available: https://www.researchgate.net/profile/Ronald-Giere/publication/238311790_Scientific_Cognition_as_Distributed_Cognition/links/556f83d508aefcb861dda85f/Scientific-Cognition-as-Distributed-Cognition.pdf

52. Palacios ER, Razi A, Parr T, Kirchhoff M, Friston K. On Markov blankets and hierarchical self-organisation. J Theor Biol. 2020;486: 110089. doi:10.1016/j.jtbi.2019.110089

53. Albarracin M, Demekas D, Ramstead MJD, Heins C. Epistemic Communities under Active Inference. Entropy . 2022;24. doi:10.3390/e24040476

54. Heins C, Klein B, Demekas D, Aguilera M, Buckley CL. Spin Glass Systems as Collective Active Inference. Active Inference. Springer Nature Switzerland; 2023. pp. 75–98. doi:10.1007/978-3-031-28719-0_6

55. Levin M. The Computational Boundary of a “Self”: Developmental Bioelectricity Drives Multicellularity and Scale-Free Cognition. Front Psychol. 2019;10. doi:10.3389/fpsyg.2019.02688

56. Levin M. Endogenous bioelectrical networks store non-genetic patterning information during development and regeneration. J Physiol. 2014;592: 2295–2305. doi:10.1113/jphysiol.2014.271940

57. Ramstead MJD, Veissière SPL, Kirmayer LJ. Cultural Affordances: Scaffolding Local Worlds Through Shared Intentionality and Regimes of Attention. Front Psychol. 2016;7: 1090. doi:10.3389/fpsyg.2016.01090

58. Vasil J, Badcock PB, Constant A, Friston K, Ramstead MJD. A World Unto Itself: Human Communication as Active Inference. Front Psychol. 2020;11: 417. doi:10.3389/fpsyg.2020.00417

59. Rouse J. Scientific Practice and the Scientific Image. 2015 [cited 10 May 2023]. doi:10.7208/chicago/9780226293707.003.0006

60. Salthe SN. Creating the Umwelt: From Chance to Choice. Biosemiotics. 2014;7: 351–359. doi:10.1007/s12304-014-9204-1

61. Ramstead MJD, Badcock PB, Friston KJ. Variational neuroethology: Answering further questions: Reply to comments on “Answering Schrödinger’s question: A free-energy formulation.” Physics of life reviews. 2018. pp. 59–66. doi:10.1016/j.plrev.2018.01.003

62. Ramstead MJD, Constant A, Badcock PB, Friston KJ. Variational ecology and the physics of sentient systems. Phys Life Rev. 2019;31: 188–205. doi:10.1016/j.plrev.2018.12.002

63. Palacios ER, Isomura T, Parr T, Friston K. The emergence of synchrony in networks of mutually inferring neurons. Sci Rep. 2019;9: 6412. doi:10.1038/s41598-019-42821-7

64. Friedman D, Applegate-Swanson S, Choudhury A, Cordes RJ, El Damaty S, Guénin—Carlut A, et al. An Active Inference Ontology for Decentralized Science: from Situated Sensemaking to the Epistemic Commons. 2022. doi:10.5281/zenodo.6320575

65. Hofstadter DR. Gödel. Escher, Bach: an etemal golden braid, New York. 1979. Available: http://www.ratiocination.org/Courses/phil3505/wp-content/uploads/2017/08/Hofstadter__Godel-Escher-Bach-excerpt-on-MU-puzzle.pdf

66. Baddeley RJ, Franks NR, Hunt ER. Optimal foraging and the information theory of gambling. J R Soc Interface. 2019;16: 20190162. doi:10.1098/rsif.2019.0162

67. Friedman DA, Tschantz A, Ramstead MJD, Friston K, Constant A. Active Inferants: An Active Inference Framework for Ant Colony Behavior. Front Behav Neurosci. 2021;15: 647732. doi:10.3389/fnbeh.2021.647732

68. Helmholtz. Treatise of physiological optics: Concerning the perceptions in general. Classics in psychology.

69. von Helmholtz H. The Facts in Perception. In: von Helmholtz H, editor. Epistemological Writings: The Paul Hertz/Moritz Schlick Centenary Edition of 1921 with Notes and Commentary by the Editors. Dordrecht: Springer Netherlands; 1977. pp. 115–185. doi:10.1007/978-94-010-1115-0_4

70. Gregory RL. Perceptions as hypotheses. Philos Trans R Soc Lond B Biol Sci. 1980;290: 181–197. doi:10.1098/rstb.1980.0090

71. Bruineberg J, Kiverstein J, Rietveld E. The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. Synthese. 2016; 1–28. doi:10.1007/s11229-016-1239-1

72. Campbell JO. The Knowing Universe. Independently Published; 2021. Available: https://books.google.com/books/about/The_Knowing_Universe.html?hl=&id=NMO1zgEACAAJ