European Journal for Philosophy of Science (2021) 11: 94
https://doi.org/10.1007/s13194-021-00408-y

PAPER IN HISTORY AND PHILOSOPHY OF SCIENCE

Aligning the free-energy principle with Peirce's logic of science and economy of research

Majid D. Beni[1] · **Ahti-Veikko Pietarinen**[2,3]

Received: 11 July 2020 / Accepted: 27 July 2021 / Published online: 10 September 2021

Abstract
The paper proposes a way to naturalise Charles S. Peirce's conception of the scientific method, which he specified in terms of abduction, deduction and induction. The focus is on the central issue of the economy of research in abduction and self-correction by error reduction in induction. We show how Peirce's logic of science receives support from modern breakthroughs in computational neuroscience, and more specifically from Karl Friston's statements of active inference and the Free Energy Principle, namely the account of how organisms' capacity to decrease the discrepancy between the expected value and actual outcomes entails the minimisation of errors in their hypotheses about the world. A scientific account of organisms' capacity to choose policies and form expectations is aligned with Peirce's theories of abduction and induction, and especially with the economy of research. The upshot is the recovery of Peirce's theory of the logic of science in the context of active inquiry.

Keywords peirce · abduction · deduction · induction · active inference · free energy principle · economy of research · policies · expectations

1 Introduction

Peirce offers a plausible theory of scientific method consisting of the three interlocking stages of reasoning types, namely abduction, deduction and induction. The two ends, abduction which suggests a hypothesis by a sort of a scientific guess, and induction which tests the models, are mediated by deduction that infers from the

✉ Ahti-Veikko Pietarinen
ahti.pietarinen@gmail.com; ahti-veikko.pietarinen@taltech.ee

Majid D. Beni
mbeni@metu.edu.tr

1 Middle East Technical University, Ankara, Turkey

2 Tallinn University of Technology, Tallinn, Estonia

3 Research University Higher School of Economics, Moscow, Russia

hypotheses classes of testable predictions. The three have specific qualities complementing each other. Abduction is the least secure but the most fertile of the three and the only one that can suggest altogether new ideas. Induction is likewise fallible but it gives stronger reasons to adopt its provisional conclusions because those conclusions have "been reached by a rule of inquiry which persistently followed will ultimately correct any error into which it may have led the inquirer" (R 905, 1907).[1] Deduction, upon which Peirce takes the two other modes of reasoning to depend is, by contrast, the most secure as it "concludes no fact not a part of the state of things that its premisses assert" (*ibid.*). These three inference's distinctive qualities are the economy of research (abduction), analysis and necessary reasoning (deduction), and self-correction by error reduction (induction).

There have been a few remarkable attempts at unpacking and defending the economic and non-epistemic aspects of Peirce's theory (Rescher, 1976, 1978; Wible, 1994). More recently, these ideas of Peirce kindled new interest (Chiffi & Pietarinen, 2019a, b; Chiffi et al., 2020; Mayo, 2005; Pietarinen, 2014, 2018). In this paper, new grounds are explored in support of Peirce's naturalistic theory of scientific reasoning, especially induction and abduction, with a focus on Peirce's theory of the economy of research. We will have more to say on deduction – the crucial link between abduction and induction and the basis of reasoning upon which the other modes repose – in another context, as that important part will not find enough room for exposition in the present paper. In brief, deduction is by no means limited to the type of necessary reasoning characteristic of mathematical demonstrations. In Peirce's theory, deduction consists of *logical analysis* (or the theory of real definitions) and *demonstrative reasoning* (R 905; Pietarinen & Bellucci, 2014). The former involves the development of logical tools, systems and diagrammatic notations to perform tasks to analyse various phenomena, while the latter concerns actually carrying out deductive reasoning steps in two distinctive fashion, namely as corollarial ('trivial') or theorematic ('creative') inferences. This latter distinction, unlike the former, has been rather extensively discussed in the literature. Moreover, reasoning about probabilities (the doctrine of chances) is also part of the deductive mode of reasoning (and not induction) in Peirce's logic of science. In the present paper, we endeavour to show that there is some remarkable intellectual affinity between these classical ideas and the recent statements of the Free Energy Principle (FEP) and its corollary theory of active inferences.[2]

Although the enterprise in this paper is basically philosophical, the grounds that we are unveiling are mainly scientific. We assume that the attempt at reinforcing our philosophical pursual with scientific grounds would be plausible both from a Peircean perspective and some new naturalist takes on philosophy (such as Beni, 2019a; Ladyman & Ross, 2007). This paper aims to account for the self-correcting

[1] References to Peirce's works are standard. R is to Peirce (1967) by the Robin number. CP is to Peirce (1931–1966) by volume and paragraph number. EP is to Peirce (1998).

[2] See Beni & Pietarinen (2021a, b) for some initial studies along these directions, incorporating what Peirce takes to be characteristics of deductive reasoning, namely model-based and diagrammatic reasoning, as a central element of organism's process of building representations of their generative models.

mechanisms of induction and the economic character of abduction that underpin scientific activity in terms of recent breakthroughs in computational neuroscience. That is to say, it grounds Peirce's logic of science, and especially abduction and induction, in a recent scientific study of some specific natural phenomena: organisms' predictive processing mechanisms, their prediction error minimisation ability, and their nearly uncanny capacity for staying in the state of homeostasis with their environment are the kinds of naturalised phenomena that Peirce specified in terms of surprises, guesses, expectations and anticipations guided by the considerations of the economy of research (abduction), as well as in terms of reliable error reduction in self-correcting processes (induction). In short, the paper draws on FEP and active inference to defend and vindicate a broadly Peircean account of scientific inquiry.

FEP lies at the nub of a viable scientific account of perception-cognition in terms of probabilistic inferences. Among others, FEP provides a general framework that unifies the Bayesian brain theory, game-theoretic account of natural selection (Maynard Smith & Price, 1973), and neuro-computational theories of action, cognition, perception, and consciousness (Friston, 2010; Friston & Frith, 2015; Friston & Stephan, 2007). FEP has been used to explain how organisms' capacity to decrease the discrepancy between the expected value and actual outcomes entails the minimisation of errors in their hypotheses about the world. Peirce wondered how our guesses at hypothetical explanations can at least sometimes turn out right, and often enough so that discovery and scientific progress ensue that would not be possible should the guesses be blind and draw upon unsystematic procedures. In the light of abductive guesses, FEP serves as the generalised guiding principle for *active inferences*, inferences that can change the organisms' *internal* states according to the expectations that the organism entertains. It also explains how multi-scale directedness of living beings leads to the choosing of such strategy profiles (general habits of action) and stable policies that through organisms' forming of new expectations allow them to optimise their cognitive grip on the world. The upshot of the present paper then is that FEP's account of prediction error minimisation, thick depth of its generative models, and policymaking support Peirce's pioneering account of the logic of scientific inquiry.

The paper is structured into the following sections. It proceeds immediately by outlining Peirce's conception of scientific inquiry with the focus on Peirce's notion of induction first (Sect. 2). Section 3 outlines FEP and explains how FEP as an emerging research program in computational neuroscience supports inductive reasoning as a real faculty of cognition. It is shown that FEP's account of prediction error minimisation provides a biological basis for the explication of inductive self-correcting processes. In Sect. 4, the paper returns to Peirce's text, outlines his notions of abduction and expectation, and revisits his theory of the economy of research. In Sect. 5, the paper moves on to show how abduction and the economy of research can be cashed out in terms of FEP. These notions are accounted for in terms of the organisms' ability to make temporally thick generative models that can accommodate expected values and future-oriented hypotheses about the world. In the coda of the paper (Sect. 6), it is explained that the Peircean conception of scientific method is realistic in the strong sense in which it is aligned with what our modern science (in terms of FEP) tells us about scientific practices as evolved forms of basic prediction error minimisation mechanisms in the natural world. We conclude the paper by submitting that the same basic mechanisms that maximise the survival of fit organisms also underpin abduction and the economy of research, and (pending the incorporation of deduction into this picture in full through model-based and diagrammatic reasoning; see Pietarinen & Beni, 2021a, b) Peirce's inferential theory of scientific inquiry in general.

2 Induction and the self-correcting thesis

Science is an agent-based and action-oriented collaborative social and ecological enterprise. Its goal-directed activity is rendered by actual human beings as cognitive agents who are situated in specific social contexts and ecological environments (Magnani, 2009, 2017). There are already theories of philosophy of science that acknowledge this point (Beni, 2019a; Giere, 1987, 2010; Weisberg, 2007). However, none of the mentioned philosophical enterprises is fortified with a viable account of the logical method of scientific inquiry.

We submit that scientific practice is underpinned by non-deductive and strategic rules of action that are ampliative, defeasible, and non-monotonic (Chiffi & Pietarinen, 2018; Gabbay & Woods, 2005, 2006; Hintikka, 2007). These rules are liable to annulment or correction, and the conclusions that are drawn from their premises are tentative and revisable. Scientific hypotheses are "accepted on probabtion" (Peirce, 1913): they are sensitive to context and the agent's or institute's goals and interests; the issue of their demarcation is not to be represented as a simple decision tree. Despite being aware of this fascinating complexity, we intend to restore a sophisticated, flexible, and scientifically defendable account of scientific methods based on Peirce's pioneering work. His account of what scientific methods are to address the question, first, in terms of the theory of abduction, and more specifically in terms of economic considerations, considerations of costs and benefits of scientific research, including the quality of *uberty* (fruitfulness grounded on truth) that some such successful scientific hypotheses are endowed with (Peirce, 1913; Pietarinen, 2020). Such economic considerations outweigh the choice of hypotheses that are worthy of being tested as candidates for good theories as well as the mechanisms of verifying those hypotheses (see Chiffi & Pietarinen, 2019a, b). Considerations of utility and cost of research can find a normative status as a non-epistemic value in such contexts (Chiffi et al., 2020).

After choosing viable hypotheses (with expected uberty), scientists subject the hypotheses to empirical tests. They strive to improve upon the formulation of fledging theories on the basis of feedback from experimentation. This is the inductive phase. Induction is a tricky notion and it could be unpacked in a number of ways. According to Deborah Mayo's reading, Peirce provides an ingenious conception of induction in terms of *reliability*, in the sense that inductive mechanisms ensure the trustworthiness of inferences on the basis of "the ability to control error probabilities of test procedures" (Mayo, 2005, p. 301). Mayo phrases Peirce's theory of trustworthy or reliable inductions in terms of self-correction:

> **Self-Correcting Thesis SCT**: methods for inductive inference in science are error correcting; the justification for inductive methods of experimental testing in science is that they are self-correcting. (Mayo, 2005, p. 299)

Mayo also argues that this understanding of scientific induction is in line with statistical approaches to error-reduction, rather than Bayesianism which may be prone to subjectivism. It could be contended that the statistical approach to error reduction is concerned with quantitative aspects of induction and as such does not consider qualitative induction, meaning that although Peirce's thesis indicates why theories approach truth in the long run, it fails to explain how the rejected theories come to be replaced by better ones (Mayo, 2005, p. 302; Rescher, 1978). Mayo convincingly argues that the objection unnecessarily construes induction in the context of orthodox theories such as Hempel's H–D model or straight rule. Mayo, on the other hand, argues that the reliability of scientific induction is due to its capacity to reduce errors or be self-correcting. Since Peirce's three stages of reasoning form an organic and interconnected whole of the logic of science, one could add that induction alone is not self-correcting, but that the full cycle of the three stages of scientific reasoning (abduction-deduction-induction) is (Pietarinen & Bellucci, 2014).

Error reduction (also self-correction or error-correction, as Mayo calls it) is an experimental process that finds clear statistical expression. It needs to satisfy two main conditions (Mayo, 2005, pp. 305–306).

1. Error reduction should allow for learning from the results of confrontation between the theory and evidence, so that the method could detect the discrepancy between predicted and actual outcomes.
2. The method should be able to check its own premises so that in the case of detecting error they (i.e., errors) could be subtracted from the methods of analysis of pieces of evidence.

On top of these two conditions, data that is going to be tested must be garnered actively, and it is here that the notions such as serendipity, preparedness and epistemic imperatives (as leading principles of reasoning) lord over scientific endeavours in general (see also Bellucci & Pietarinen, 2020a, b; Chiffi & Pietarinen, 2019a, b; Pietarinen & Bellucci, 2014; Pietarinen & Beni 2021b). Moreover, the process of error reduction may reduce the discrepancy between predictions and observations but need not eliminate predicted errors completely. At any rate, induction consists in an error-correcting or error reducing mechanism. Next, we describe the FEP and show how it could be used to naturalise the main aspects of Peirce's account of scientific reasoning.

3 The free energy principle and the error-correcting mechanisms of inductions

According to Mayo (2005), Peirce's account of induction was based on statistical self-correcting mechanisms whose reliability could be further substantiated. In this section, we will unfold the cognitive basis of induction and error-correcting mechanisms and account for their reliability in scientific terms.

There is a flourishing framework of computational neuroscience that can provide a viable (and almost objective) reliable error-reducing mechanisms of cognition. We outline this theory and exploit it to argue that the most conspicuous virtue of Peirce's account of the scientific method is its agreement with the facts that a state-of-the-art theory of cognitive science can reveal about how human beings draw models of the causal structure of the world. Below, we will flesh out the key requisite details.

We (as cognitive agents) generate models of (or hypotheses about) the causal structure of the world, and use these generative models to represent the causes of our sensations through statistical inferences. Historically, this probabilistic-inferential approach to perception was introduced to neurophysiology by Hermann von Helmholtz in the nineteenth century. More recently, Karl Friston and his collaborators have developed such early theories on numerous occasions and rephrased them under the rubric of the Free Energy Principle (FEP). The status of FEP as a theory of cognitive science has been controversial, and some (see e.g. Colombo & Hartmann, 2015; Colombo & Wright, 2016) challenge its theoretic status while others (Beni, 2018a, 2019b) defend it. What is less controversial is that FEP provides a unifying framework that subsumes various streams of research in cognitive science. It comprises Bayesian theories of cognition (Daw et al., 2005; Dayan et al., 1995), predictive processing theory (Rao & Ballard, 1999) and optimal decision theory (Knill & Pouget, 2004) (for a review of the history see Friston, 2012a). FEP unifies these theories of optimisation and supplements them with an evolutionary account of the organisms' dynamical interactions with their environment. Organisms do not have direct access to the causal structure of the world and could represent it through top-down mechanisms of probabilistic active inferences.

To preserve the state of homeostasis, organisms must be able to optimise their probabilistic inferences and their models/hypotheses. FEP's framework provides a basis for accounting for the organism's optimal capacity for representing the world's causal structure on an inferential-probabilistic basis. Integral to FEP's account of the agent-world dynamical relationship is the notion of "active inference". The notions of surprise, active inference, and its top-down mechanisms of information processing now provide a scientific framework for grounding Peirce's theory of scientific inquiry. Also, although FEP is related to the family of Bayesian theories, its formal framework could be construed either in terms of statistical frequencies or incorporated into an entropy-based framework altogether (see Beni, 2018b for details). The latter would ease the friction between having to choose either frequentism or Bayesianing as the foundation of probabilities in this particular argument.[3]

[3] This is also in the (implicitly) Peircean spirit of Jaynes (1995/2003: *xii*) who writes: "[N]either the Bayesian nor the frequentist approach is universally applicable, so in the present more general work we take a broader view of things. Our theme is simply: *Probability Theory as Extended Logic*. The "new"

Free energy is an information-theoretic measure (Friston, 2010, p. 2). Assuming that $s\sim$ is the set of sensory signals, $p(\tilde{s}, \vartheta|m)$ is the probabilistic (generative) model m that generates sensory consequences $s\sim$ from their causes ϑ, and $q(\vartheta)$ is recognition density or the probabilistic representation of environmental causes. With these constructs in place, we can define the free energy in the following way.

$$F = -\langle lnp(s \backsim, \vartheta|m)\rangle_q + \langle lnq(\vartheta)\rangle_q$$

which, in case that perception optimises prediction, can be states as

$$F = D(q(\vartheta|\mu)||p(\vartheta|s \backsim)) - lnp(s \backsim |m)(assuming\mu = argmax\,Divergence)$$

This expression shows that free energy is a bound on the negative log evidence or surprisal -ln $p(s\sim|m)$. Minimising free energy thus entails the principle of maximum entropy, according to which amongst all posterior probabilities of a state the distribution with the largest entropy <-ln $q(\vartheta)$> q represents the state best (Friston, 2013). This should be contrasted with the entropy of sensory data, which is the expected surprisal <-ln $p(s\sim|m)$>. Generally, the entropy of sensory states tends to increase, but in order to resist the dispersing effect of the environment (i.e., the second law of thermodynamics), organisms can place an upper bound on their sensory entropy by minimising free energy. Organisms that keep their free energy low are, on average, found in the sensory states that are consistent with their generative model m; that is, with the kinds of sensations they expect to encounter. As this formulation indicates, mechanisms of hypothesis-making that FEP describes are based on top-down predictions: hypothesis-making processes begin with internal (generative) models within the organism's (embodied, cognitive) system. It is only after constructing the more refined classes of hypotheses with the aid of the intermediary of deductive abilities that the agent tries to test her predictions by seeking and finding confirming evidence through sampling the environment. A generative model is defined in terms of the joint probability over causes and data. It comprises likelihoods, which are the probabilities of sensory data, given the causes of sensation (Friston, 2010, p. 3), and prior beliefs over the causes. The inversion of a generative model corresponds to *active inference* based on predictions about the sensed world. Accordingly, perception is a process of reducing prediction error, in the sense that perception consists in "inverting the likelihood model (mapping from causes to sensations) to access the posterior probability of the causes, given sensory data (mapping from sensations to causes)" (Friston, 2010, p. 3). Hypotheses that the organism entertains about the world generate predictions about the results of the agent's encounter with her world. Generating plausible predictions by analysis, articulation and refinement comes round to the natural correlate to the deductive stage in Peirce's logic of scientific inquiry, mapping forms of the model into other forms by a series of transformations

Footnote 3 (continued)

perception amounts to the recognition that the mathematical rules of probability theory are not merely rules for calculating frequencies of "random variables"; they are also the unique consistent rules for conducting inference (i.e. plausible reasoning) of any kind". Jaynes could almost have inserted the term "abduction" within the parentheses.

on representations of the model so that the predictive material is put under the clearest light in the conclusion. The organism then endeavours to sample data from the environment in order to confirm her hypotheses. This, in turn, is to make (weak) inductions. Sensory inputs, as pieces of evidence, get tested against the predictions. Perception, therefore, results from a comparison between the top-down flow of predictions and the upward going flow of evidence. The brain (or the organism's body plan overall) uses the discrepancy between predicted input and evidence (i.e., surprisal) to update beliefs about the causes of perceptions that are beyond the evidentiary boundary (Hohwy, 2014). This, in turn, involves abduction (Pietarinen & Beni, 2021a, b), the initial stage in the logic of science. Perception comprises error-correction which is the same thing as error-reduction in the sense that the organism revokes its primary hypothesis about the world or revises it on the basis of the feedback received from the comparison between predicted inputs and actual inputs. This is sometimes referred to as self evidencing.

Returning briefly to Mayo's construal of Peirce's account of induction, it can be assumed that it is SCT, the self-correction thesis, which is the cornerstone of Peirce's theory of induction. Mayo also proposed that self-correcting mechanisms should satisfy the two conditions (1)–(2) stated above. This means that the mechanisms should allow learning from the results of confrontation between the theory and evidence so that the method could detect the discrepancy between predicted outcomes and actual outcomes. They should also be able to check their own premises so that in the case of detecting error those errors could be subtracted from the methods of analysis of pieces of evidence. Mayo also submitted that evidence and data that confirm hypotheses should be randomly garnered. As the broad-brush outline that we have depicted above indicates, all of these factors are present in FEP and its account of prediction error minimisation. We now proceed to a specific account of the fit between Peirce's account of induction and FEP.

SCT can be cashed out in terms of Prediction Error Minimisation (PEM) which is inbuilt into cognizer's systemic structures. PEM provides a basis for (some) organisms' capacity to be able to adjust their hypotheses about their environment based on the feedback that they receive when they compare their expectations with their environment and surmise that there are discrepancies and errors. This refers to the first condition that Mayo associated with SCT. Indeed the scientific method should be sensitive to discrepancies and errors. FEP could explain how this condition could be satisfied in the natural world, namely based on top-down prediction error minimisation mechanisms that are arranged hierarchically. This means that generative models of causes of sensation at each level of the system generate causes at the lower level, down to the sensory data at the lowest level (Friston, 2010, p. 3). The discrepancy between generative models and sensory data indicates that the generated models of causes were not precise enough. FEP also accounts for the satisfaction of the second condition, explaining how it is that some organisms could revise their premises with reference to the hierarchical mechanisms of PEM. After spotting the discrepancy between the predicted inputs and the actual ones, the system uses this discrepancy (i.e., free energy or surprisal) and updates models of causes at each level up to the prior models of the causes at the topmost level. Reducing error and minimising free energy is tantamount to one another, and a given organism could minimise its free energy not only by passively updating its generative models but also by action via intervening in the environment as well as sampling evidence randomly. The update of the priors (or revising premises, according to Mayo's vocabulary) is partly the result of the environment's enforcement of sensory impressions on the agent's cognitive system, while the agent too could change the states of the environment. Action can reduce free energy by intervening in causes of sensations in the external world and thus manipulating sensory inputs (Friston, 2009, p. 295). For example, if I find myself looking at the sun, I can either note – with some surprise – "I am the sort of creature that can look at the Sun". Or, alternatively, I will immediately look away, fulfilling my prior beliefs that "I am not sort of creature that looks at the sun." Clearly, if I am not the sort of creature that can look at the sun, actively sampling the right kind of visual sensations will underwrite my adaptive fitness.

It is worth emphasising that FEP characterises actions as evidentiary inference, by associating them with *policies*. In other words, policies are ordered sequences of actions selected to minimise free energy expected after following a course of action (Hohwy, 2014, p. 271). Agent's perceptual mechanisms are alloyed with action to reduce surprises on the basis of their general policy for maximising survival. This means that "action and perception are integral parts of the same inferential process and one only makes sense in light of the other" (Friston et al., 2013, p. 2). To minimise their entropy (i.e., the average of surprisal or uncertainty), the organisms endeavour to intervene with the causes of their sensory inputs and shape them according to their internal models of the world. Thus, the second condition of SCT can be explicated based on FEP too.

Finally, we remember that Mayo suggested that induction should be understood in terms of "reliability" and "trustworthiness". FEP can be used to account for the reliability of induction too. But in the Peircean framework, the reliability of induction is to some extent connected with the antedating economic considerations that prevail in abduction. Thus, we can account for the reliability of the scientific method in this sense (Sect. 5) and then explain how FEP accounts for the trustworthiness and reliability of PEM (or SCT).

Similar considerations underpin accounts of the cognitive basis of abduction too and are taken up in Sect. 4 next. For the time being, suffice it to say that the problem of reliability in FEP's framework can be addressed in terms of the role of the success of the organism's hypotheses in *maximising* the survival of the organism. Inductive mechanisms of organisms that minimise their prediction error successfully are reliable in the sense that they can get updated economically. In this respect, FEP draws its force from its connection with theoretical biology. There are evolutionary reasons to explain why agents need to be able to optimise their models of the world. Metaphorically speaking, natural selection favours organisms whose neurocognitive systems are developed in a way that enables them to optimise their prior expectations objectively, namely, by updating their internal models of hidden causes on the basis of sensory evidence. Hence error reduction mechanisms are reliable. Next, we shall proceed to abduction in the context of Peirce's philosophy.

4 Abduction, expectations and the economy of research

Peirce's account of induction is linked with his theory of abduction and serves as a partial justification of the latter (Bellucci & Pietarinen, 2020a, b). In brief, abduction concerns "the process of forming explanatory hypotheses" and is "the only logical operation which introduces any new idea" (CP 5.171, 1903). As Peirce discusses the novelty of abduction in connection to his philosophy of pragmati(ci)sm, his identification of the notion in these 1903 Harvard lectures develops in his later years into a promising account of future-oriented *expectations*. Wible paraphrases this point by remarking that, "Peirce's theory of expectations is embedded in his conception of mathematics and cognition in an evolutionary world." (Wible, 2019, p. 20).

It is important to be clear about the meaning of Peircean expectations. Expectations are results of general, stable and enduring *habits* that organisms must take and possess in order to survive in a world that offers specific opportunities and affordances for survival.[4] Expectations concern the possibility of an agent's conjectures turning out true. They concern what Peirce termed "possible and esperable uberty" of hypotheses that are conjectured. The "esperable uberty" is the quality of those hypotheses that are endowed with some "*value* in productiveness" (Peirce, 1913). Such values are not only epistemic values. Uberous hypotheses are those that in Peirce's terms are "*gravid* with young truth" (*ibid.*). He is not talking plainly about the truth and falsity of hypotheses. By "young truth" he appears to mean the values that scientists consider hypotheses to have as soon as their further analyses necessitate the application of also some *non-epistemic* values such as considerations of economy. Moreover, such general suggestions at the modalities of what "may-be" need not be grounded on fully articulate, intersubjectively contestable reasons. Modifications of internal data could result in the formation of such expectations. Yet Peircean expectations are not singular states of mind. Nor are they concerned with what individual decision-makers happen to contemplate in any given epoch of time, given that expectations on what would follow from adopting certain hypotheses on probation have some important collective, communal and social dimensions.[5]

Although we can thus find support to Peirce's theory of science in the findings of contemporary computational neuroscience, the ultimate meaning of expectations is not intended to be psychologistic, that is, determined by theories of psychology. Likewise, in order to be able to produce novel fruitful scientific theories that can accommodate new predictions, we must be able to specify new patterns of relations between classes of phenomena. Scientists demarcate some hypotheses as worthy

[4] This technical meaning of expectations is different from its common meaning, which Peirce bundled with notions such as desires, wishes and intentions. The latter are not general otherwise than through connection with a concept. These notions do not give"ultimate intellectual interpretants" (R 318; EP 2: 430, 1907) and thus not the meaning of signs, as the meaning that they have has to be effected by reference to concepts. Expectations *in futuro*, in turn, appear in "mental forms" (CP 2.86) as intentions do, but do not concern conceptual meaning but the "mental effects of the signs that they interpret" (R 318; EP 2: 431). Such expectations are given in conditional forms (subjunctives), while singular expectations as such are merely non-conditional judgments.

[5] These dimensions are sketched somewhat further in Beni & Pietarinen (2021a).

of further testing and scrutiny under fundamental uncertainty, on the basis of their expectations about novel experimental results of those hypotheses in the future. What matters as to the value of such hypotheses is both the *interconnectedness of evidence*, often sparse, between the number of guesses that had to be made during their generation, and the *expected novelty and gain*, should those hypotheses be on the right track. Attempts at reducing uncertainty about the future outcomes of frontier sciences and discovery are explicated by Peirce under the rubric of abduction. Indeed abduction is the type of reasoning "that creates hypotheses to account for the surprising facts by guessing" (Chiffi et al., 2020, p. 2; Pietarinen, 2014). Scientific guesses are particular kinds of sampling processes concerning open-ended future states of affairs.

The meaning of expectation can now be cashed out in terms of the notion of *surprise*, which Peirce (1905) explicitly took on board only after his more famous but merely promissory Harvard scheme of abduction, when he described the abduction as "Reasoning from Surprise to Inquiry": given a (surprising) fact *C*, if *A* implies *C*, then *it is to be inquired whether A* plausibly holds. Like abduction elsewhere in his writings, this later schema begins with an observation of a surprising fact or event, and through a conditional major premise concludes, in a 'co-hortative' mood which Peirce termed "the investigand" (Ma & Pietarinen, 2018), that something ought to be done. By asserting conjectural conclusions in such investigand moods, those conjectures become the ratified subjects of inquiry: the purpose of future inuiry is then to find out whether they plausibly are the matters of course.

Peirce wanted to, first, characterise abduction in ways that would avoid earlier conflations of it with induction (and with IBE, see e.g. Seth, 2014 for an example of such conflation), and to provide a more nuanced expression of its logical, linguistic and semiotic forms than was possible with the rough sketch given of it in his 1903 Harvard Lectures. For one, Peirce added the investigand mood to qualify how abduction's conclusion is to be asserted as a matter of fact.

These points aggregate into a preliminary basis upon which one could begin to discern how FEP can contribute to naturalising abduction. Free Energy itself is a mathematical bound on surprisal on sampling some data, and surprisal is the negative log-probability of an outcome (Friston, 2010, p. 2). We will elaborate on this point in the next section, but for the time being, suffice it to state that because scientists' predictions of the future results of given hypotheses are based on their conjectural expectations, expectations about the future turns of events have a central status in the context of Peirce's theory.

Peirce's theory of abduction, in turn, and supplemented with his theory of economy of research (Peirce, 1876), provides a forward-looking theory of demarcating hypotheses that could (*in futuro*) lead to the discovery of real patterns or at least underlie promising streams of inquiry that are not altogether dislodged from the reality of the patterns in the world. Time, resources and energy prevent all hypotheses to be submitted for further verification and one has to choose hypotheses that are put in our docket based on considerations of the economy such as costs and benefits that override all other considerations (CP 5.602). Our expectations of what happens in the future provide a cognitive basis for our calculation of costs and benefits. Error correction mechanisms are associated with the mechanisms of projecting expectations onto the future. Error correction consists of revising inaccurate expectations and increasing their precision based on the feedback from the encounter with observations.

What Peirce states, already a year before the 1903 Harvard lectures, was:

> Looking at an expectation referring to the distant future, something, for example, which you propose to do, it is very little distinguished from a mere “may be.” There is a sort of picture in your imagination whose outlines are vague and fluid. You do not attach it to any definite occasion, but you think vaguely that some definite occasion there is, to which that picture does attach itself, and in which it is to become individualized. (CP 2.146, “Why Study Logic?”, 1902)

This quotation epitomises many of Peirce’s most important views on abduction that he in the coming years would much refine, including a qualification to what these “*may-be*s” are in terms of *invitations to inquiry* (Ma & Pietarinen, 2018; Pietarinen & Bellucci, 2014). There is also an important iconic and imaginative character to abductions to be noted (Bellucci & Pietarinen, 2016, 2020a, b; Champagne and Pietarinen, 2018). In the present paper, however, another important implication of Peirce’s remarks is advanced by resources made available by recent theoretical and experimental psychology, especially in view of his continuing interest in the anatomy of the brain and the early neurosciences (Cristalli & Pietarinen, 2020; Pietarinen, 2006). Surely Peirce’s approach not only allows for but recommends drawing on science itself in support of his philosophical ideas. In fact, Peirce’s philosophy of science is so scientific-minded[6] that claiming it to be based on thoughtful speculation would be seriously off the mark. Peirce’s account of the scientific method in terms of abduction and the economy of research is plausible not only on such philosophical grounds but also because scientific facts support and validate it. In the next section, we endeavour to provide some such essential ingredients for a scientific basis for naturalising Peirce’s theory of the economy of research and abduction.

5 Free energy principle, expectations and abduction

Recently, naturalising abduction has been advanced in several works, including Magnani (2009) and Park (2017). A general remark on the possibility of drawing on the resources of modern psychology in support of Peirce’s theory of abduction has been made in Chiffi et al. (2020). Boltzmann machines are examples of neural nets that perform inverse reasoning from effects to causes. Moreover, Murray Shanahan had explained how an account of abductive perception could be developed in the

[6] Among other places, this is evident in Rescher’s interpretation of Peirce’s take on science, according to which, “Science is autonomous. Corrections to science must come from science... The mistaken results of science can be improved or corrected only by further results of science. There can be no recourse at this point to tealeaf reading, numerology, the Delphic oracle or the like” (Rescher, 1978, p. 160).

context of robotics and AI with an eye to theoretical notions of top-down information processing and active inferences under uncertain circumstances. In developing his account of expectation as the basis of top-down mechanisms, Shanahan submits that despite possible consequences, to the extent that his enquiry is at issue, "the aim is not necessarily to achieve psychological or biological plausibility but to produce a well-engineered mechanism" (Shanahan, 2005, p. 108).

The objective of this paper is to provide an alternative, moderately psychological and biological account of those top-down information-processing mechanisms that underpin abduction in human cognition in general and in scientific methodology in particular. In the previous section, it was explained that Peirce's theory of the scientific method is importantly characterised by his central notion of scientific *expectation*. According to Peirce, "Our knowledge of any subject never goes beyond collecting observations and forming some half-conscious expectations, until we find ourselves confronted with some experience contrary to those expectations" (CP 7.36). Wible clarifies: "The abduction could start with some awareness of a surprise or an unexpected anomaly. An abduction is the idea that might make sense of that surprise. As such an abduction amounts to a hypothesis that is worthy of investigation" (Wible, 2019, p. 25). FEP, we go on to propose, provides a theoretical handle on naturalising expectations and abduction.

As the overview of FEP in Sect. 3 indicated, organisms with objectively revised prior expectations can maximise their survival because, first, they can invest in reliable hypotheses about their environment, and second, they can adjust their hypothesis in light of surprise or prediction error. As explained, the agent acts on the basis of those expectations. As a result of the agent's action, the environment changes and furnishes new posterior probability distributions that provide new evidence for the agent's models. The agent samples from the new observations and update its internal models and acts again. Free energy is the bound or limit on the agent's entropy or surprisal, meaning that for the agent to survive, the divergence between its expected values and control states (as parts of generative models that refer to approximate Bayesian posteriors) shall not exceed appropriate bounds on free energy. In order to increase their chance of nourishment and prosperity and avoid dangers that lead to the extinction of their species, organisms must be able to form hypotheses that lead to reliable expectations. For instance, to maximise their survival, a fish must expect to remain in the water and a camel must expect survival in the desert, everything else being equal. A fish that expects to survive in the desert or a camel that expects to remain alive under the waves will minimise their survival (Friston, 2012b). Unrealistic expectations do not contribute to the organism's overall fitness and its survival. Mere judgmental (non-conditional) expectations won't suffice, either. There needs to be experimentation that evaluates real expectations, guesses upon what could be happening next, and those are expressed as hypothetical, defeasible forms of conditionals.

This speaks directly to Peirce's conception of expectations and their optimisation. According to Peirce's nearly final formulations of abduction,

> Every inquiry whatsoever takes its rise in the observation, in one or another of the three Universes, of some surprising phenomenon, some experience which

> either disappoints an expectation, or breaks in upon some habit of expectation of the *inquisiturus;* and each apparent exception to this rule only confirms it. There are obvious distinctions between the objects of surprise in different cases; but […] inquiry begins with pondering these phenomena in all their aspects, in the search of some point of view whence the wonder shall be resolved. At length a conjecture arises that furnishes a possible Explanation, — by which I mean a syllogism exhibiting the surprising fact as necessarily consequent upon the circumstances of its occurrence together with the truth of the credible conjecture, as premisses. (EP 2: 440-1, CP 6.469, 1908)

Of course, the form of abduction that is liable to logical characterisation is far more advanced than the primary mechanisms of expectation formation and habit taking. However, the essential capacity lies there at that basic level. Peirce, too, agrees that the formation of reliable systems of belief demands to avoid the disappointment of expectations; that is, to avoid surprises.

When faced with surprises, revising beliefs becomes a costly business. Peirce agrees that when forming these inferences, the organism favours its internal hypotheses and tries to garner confirming evidence for them (CP 6.469, 1908).[7]In any case, FEP provides a viable account of how biological systems in the real world manage to minimise their surprisal or prediction error, with an equal effort as that which has to do with maximising their expectations.

Organisms that have the capacity to invest in primary models (or hypotheses) of the environment and revise their hypotheses efficiently maximise their survival. Such organisms succeed in optimising their expected value. The expected value can be characterised in terms of the utility function of the organisms' states on the basis of evidence that the organisms can garner. In order to maximise their fitness, organisms must be able to optimise their expectations of the world. Expectations find a clear technical definition here, given that fit organisms must maximise their expected value expediently and economically. In this context, "perfect rationality corresponds to choosing states from a low temperature distribution, whose probability mass is concentrated over the state with the highest value" (Friston et al., 2013, p. 2). Here, organisms' capacity for updating the priors underpins the epistemic (or intrinsic) value of their expectations (identified with the decrease in uncertainty) (see Friston et al., 2015). This provides a scientific basis for explicating Peirce's notions of expectation and abduction. To be more precise, FEP's account of expected values are ingredients of the cognitive psychology of expectation which provides a basis for Peirce's future-looking theory of abduction, which he began to sketch in his 1902 "Why Study Logic?" in terms of "general cognitions of potentialities *in futuro*" which, "if duly constructed, will under imaginary conditions determine *schemata* or imaginary skeleton diagrams with which percepts will accord when the real conditions accord with those imaginary conditions" (CP 2.148).

[7] Such favouritism may explain some of the nearly automatic attunement of the human mind to the cognitive biases of confirmation, optimism, and the like; a topic addressed in another paper (Bobrova and Pietarinen, 2020).

While what our knowledge of such conditional expectations determines as such "skeleton diagrams" can be given rather precise logical explications (Pietarinen & Beni, 2021a, b; Gaultier, 2017; Pietarinen, 2015), the notion of expectations is worthy of additional significance and meaning drawn from what the guiding principles of such cognitive reasoning processes consist of. FEP's account of optimisation of expected value is realistic in the sense that it takes into account the cost of information processing that brings about maximisation of the expected value of survival. In this sense, FEP also provides a basis for naturalising Peirce's account of the economy of research, amounting to cost-utility calculation. Action (as an inferential process) can be modelled in terms of cost-utility decision problems in biological contexts that assign survival value to optimised beliefs about the causal structure of the world. Here too organisms have no direct access to the casual structure of the world, as the vital information of how to reach global minima are hidden from the organisms' model of the landscape. They can only form hypotheses about the causes of sensations indirectly, through the experience already encoded into their generative models, and by experimenting upon the forms those models possess, leading to informed guesses about what is going to happen next so that, at least once in a while and sufficiently often, they are tipped off onto the right track. The inferential mechanisms that are described by FEP are future-oriented in the sense that the agent builds upon evidence for its generative models in order to minimise uncertainty about causes that have valuable outcomes in the future. To maximise its survival, the organism must minimise its entropy and maximise its expected value by finding evidence for its hypotheses about the world. Not all hypotheses about the world are worth taking seriously, and expending a 'full cost model' of time and energy to check upon their validity status would be an existential mistake. The discrepancy between some of the hypotheses that the organisms make and the world is too great and the complexity of some others too high to allow for efficient revision of such hypotheses. Such hypotheses are to be revoked and replaced by those that can be revised on the basis of mechanisms of Prediction Error Minimisation and FEP. Organisms (or their sub-units) that do embody such conjecture-making mechanisms maximise their fitness and will survive. This provides a scientific basis for grounding Peirce's account of abduction and expectation in nature at multi-scale levels.

6 Temporally thick models, policies and the economy of research

In this last section, some supplementary notes are provided on the technical aspects of the FEP-based account of goal-directedness of actions and future-orientedness of expectations as a basis for naturalising abduction and expectation.

The expectations of the organisms whose course of action can minimise their free energy are related in the right way to their environment, and the optimisation of their cost-utility matrix is based on objective facts that provide a basis for their expectations in facing states of the world in future. It is this foresight into *expected free energy* that allows the agents to decrease the discrepancy between their generative models and hidden causes of sensation efficiently. Successful action in the environment is defined in terms of optimal behaviour which is dependent on the success of the agent in dissolving uncertainty by minimising expected free energy (Friston et al., 2015, pp. 201–204). Here, the important belief is that organisms not only could form hypotheses about the world, but at least some organisms can also model their expectations about the world to themselves based on their hypotheses about the future states of the world (Friston, 2018, p. 579; Friston et al., 2015). Such models are named *temporally thick models* in Friston's theory, and the key point here is that some organisms' active inferences and active sampling of data are based on the distribution of posterior beliefs over action, under the prior belief that action will minimise free energy in the future and not only in present (Friston, 2018, p. 279). By temporally thick models, one can account for the organism's agentive and purposeful behaviour (Friston, 2018, p. 579; Friston et al., 2015). This provides the basic elements needed for naturalising the Peircean account of expectation and the economy of research.

Technically, the notion of 'economy' figures prominently in the minimisation of variational free energy, in FEP's account of future-oriented mechanisms of demarcating effective strategies or policies from inefficient ones, given that some organisms possess the capacity to infer and select the policy that, if consistently pursued, will contribute to minimising the expected free energy effectively (Chen et al., 2019; Friston et al., 2015; Ramstead et al., 2019).

The story proceeds as follows: the imperative to self-evidence is met by inference and learning to minimise free energy, as specified previously in this paper. Note that the log evidence and (negative) surprisal are the same thing. This means that minimising surprise is equivalent to maximising model evidence. Crucially, log evidence can always be decomposed into *accuracy* minus *complexity*. This means that minimising surprise is equivalent to providing an accurate account of the sensorium that minimises complexity. Interestingly, acting to minimise expected surprise can also be associated with minimising expected complexity (known as *risk* in economics and Bayesian statistics), while at the same time minimising expected inaccuracy (also known as *ambiguity*). In other words, when minimising expected free energy there are epistemic and pragmatic components to any choice; namely, ambiguity and risk. The epistemic part means that much of our behaviour is about seeking information and reducing uncertainty; that is, reducing instances of expected surprise (Hohwy, 2014; Friston et al., 2013). Risk is particularly interesting, given that it refers to the expected complexity, which scores the computational complexity cost of any model or hypothesis. In this sense, an 'economic' sampling of the environment is exactly the minimisation of expected complexity cost.

Peirce proposed the qualities of incomplexity (simplicity), breadth and caution as the leading characteristics of the economy of scientific hypothesis-making. The general question then is that given a set of prima facie plausible hypotheses produced in a given research project, and given the capita that can be invested in their futher pursuit (fiscal, cognitive or otherwise), which of the many projected hypotheses should disqualify from being submitted to severe test? To mitigate risks, the quality of *incomplexity* (absence of complexity, simplicity) recommends that because few hypotheses are expected to be optimal in any case, they should at least "give a good leave" (EP 2:110): any promising hypothesis slated to be refuted should at least set an example of a good conduct to be followed, by attempting as large a 'break' as possible from it, as that would increase the possibility of discovering new hypotheses from it. Breadth gestures at unification: those hypotheses are to be favoured that are extendible in the sense of covering more ground under slightly different perspectives or adjusted boundary conditions – or can explain the unexpected in some new domains, as may happen in cross-disciplinary contexts. Caution is the economic quality of avoiding diminishing marginal returns by skilfully breaking down (by choosing a suitable method of analysis and approach to interrogation) big questions into series of small questions that can be adequately answered by the investment at hand that the conditions of the research proposal permit. For example, if questions are broken down to binary *yes–no* questions (such as to an exact definition of the null hypothesis), one is at once effectively and logarithmically reducing the size of the search space.

Of course, not all policies that the organism can pursue would maximise its survival, just as not all hypotheses will lead to fruitful scientific theories. Thus, organisms must be able to choose the right policies that provide an adequate future-oriented basis for reliable action-guiding beliefs of the organism. Policies that could maximise the organism's survival are based on optimal error minimisation mechanisms in the future (see Ramstead et al., 2019, p. 7). Thus the evolutionary basis of optimisation of the organism's expected value of policies contributes to naturalising Peirce's conceptions of abduction, expectation, and the economy of research. This is because – to state the situation somewhat metaphorically – natural selection favours organisms that pursue strategies that lead to the minimisation of the expected free energy effectively over organisms that fail to do so. Such organisms make optimal and precise models of the future states of affairs of the world based on their expected free energy. The models are optimal in the sense that they could decrease the discrepancy between expected values and control states effectively. Fit organisms are those that rely on strategies that successfully minimise the discrepancy between generative models' prediction and hidden causes of sensory data. This optimises the future-oriented inferences or expectations of such organisms and decreases their uncertainty. The reliability of the organism's probabilistic inferences and the choice of policies that maximise survival rely on one another, as "inferences about policies depend upon inferences about hidden states and vice versa" (Friston et al., 2013, p. 2). Evolutionary considerations lead to a realist understanding of the reliability of the strategies that an organism pursues to maximise its survival effectively. In this sense, future-oriented inferences of organisms that can maximise their survival are anchored to the real world.

7 Conclusions

The paper drew on a free energy account of neuro-computational mechanisms of top-down processing and active inferences in biological cognitive agents to fortify Peirce's theory of abduction and expectations. The result is a revisionary reworking of the late nineteenth century conception of the scientific method as supported by recent science. That being the case, what makes this enterprise worthwhile is the compatibility of the scientific method (in terms of induction, abduction, expectation, and the economy of research) with what our best scientific theories can tell us about the actual mechanisms of cognition, namely active inferential mechanisms that agents use to self-represent the causal structure of the world. This naturalist approach to the scientific method is not only compatible with Peirce's conception of scientific activity but further motivates and explains its essential ingredients.

In a wider perspective, FEP is not (exclusively) about human beings or their collaborations that lead to the formation of scientific theories. FEP and active inferences apply to a wide range of self-organising systems that tend to resist the dispersing effect of their environment by minimising the overall amount of their entropy. Some of such organisms happen to possess the wonderful capacity to make temporally thick models of their future states and be conscious of their insights into their expectations and policies that subsume them. In order to engage in active inferences, the organism must be able to model the consequences of its action to itself (given that the consequences postdate the action). Models of organisms that can regiment the consequences of the actions are temporally thick, where the temporal thickness of models is associated with the purposefulness and agency of the organisms. Thick models allow for the counterfactual formulation of expectations and take into account both the strategic decisions to act (namely situations that are actually never reached) and the defeasible nature of their actions (namely those represented by counterfactuals that have cancellable antecedents). Organisms with temporally thick models can model predictive posteriors that include beliefs about future outcomes, counterfactual outcomes, and hidden states, whereas the current posterior of the other kinds of complex systems covers only hidden states (Friston et al., 2015, p. 192). Some of the organisms that draft temporally thick models or working hypotheses could indeed construct exceedingly sophisticated hypotheses about their environment. They can demarcate promising hypotheses from barren ones. Promising hypotheses will be developed into flourishing theories in due course, whereas the barren ones will be left aside with memory traces on their failures imprinted on their information depositories. The choice is between those that carry what Peirce termes the 'uberty' of hypotheses and those that do not. In brief, FEP's account of prediction error minimisation, thick depth of its generative models, and policymaking have vindicated Peirce's logic of scientific inquiry sketched over a century ago.

Acknowledgements We thank Karl J. Friston for his constructive comments on an earlier version of this paper.

Declarations

Ethical approval No experimental human subjects involved.

Informed consent No experimental human subjects involved.

Conflict of Interest The authors declare no conflict of interest.

References

Bellucci, F., & Pietarinen, A.-V. (2016). The iconic moment: Towards a theory of diagrammatic imagination. In J. Redmond (Ed.), *Epistemology, knowledge and the impact of interaction* (pp. 463–481). Springer.

Bellucci, F., Pietarinen, A.-V. (2020a). Peirce on the justification of abduction. *Studies in History and Philosophy of Science A, 84*, 12–19. https://doi.org/10.1016/j.shpsa.2020.04.003

Bellucci, F., & Pietarinen, A.-V. (2020b). Icons, interrogations, and graphs: On Peirce's integrated notion of abduction. *Transactions of the Charles S. Peirce Society, 56*(1), 43–61.

Beni, M. D. (2018a). The reward of unification: A realist reading of the predictive processing theory. *New Ideas in Psychology, 48*, 21–26. https://doi.org/10.1016/j.newideapsych.2017.10.001

Beni, M. D. (2018b). Syntactical informational structural realism. *Minds and Machines, 28*(4), 623–643. https://doi.org/10.1007/s11023-018-9463-8

Beni, M. D. (2019a). *Cognitive structural realism: A radical solution to the problem of scientific representation*. Springer.

Beni, M. D. (2019). Conjuring cognitive structures: Towards a unified model of cognition. In A. Nepomuceno-Fernández, L. Magnani, F. Salguero-Lamillar, C. Barés-Gómez, & M. Fontaine (Eds.), *Model-based reasoning in science and technology. MBR 2018* (pp. 153–172). Springer. https://doi.org/10.1007/978-3-030-32722-4_10

Pietarinen, A.-V., & Beni, M. D. (2021a). Active inference and abduction. *Biosemiotics*, 1–19. https://doi.org/10.1007/s12304-021-09432-0

Pietarinen, A.-V & Beni, M. D. (2021b). Reinvigorating a 19th-century conception of the scientific method: A Peirce-pective on science. Submitted.

Bobrova, A. & Pietarinen, A.-V. (2020). Two implications and dual-process theories of reasoning. In Pietarinen A.-V., et al. (Eds.), *Diagrammatic representation and inference. Lecture notes in computer science* 12169. Springer. https://doi.org/10.1007/978-3-030-54249-8_19

Champagne, M., & Pietarinen, A.-V. (2018). Why images cannot be arguments, but moving ones might. *Argumentation, 34*, 207–236. https://doi.org/10.1007/s10503-019-09484-0

Chen, A. G., Benrimoh, D., Parr, T., & Friston, K. J. (2019). A Bayesian account of generalist and specialist formation under the active inference framework. *BioRxiv*, 644807. https://doi.org/10.1101/644807.

Chiffi, D., & Pietarinen, A.-V. (2018). Abductive inference within a pragmatic framework. *Synthese, 197*, 2507–2523. https://doi.org/10.1007/s11229-018-1824-6

Chiffi, D., & Pietarinen, A.-V. (2019a). Risk and values in science: A Peircean view. *Axiomathes, 29*(4), 329–346. https://doi.org/10.1007/s10516-019-09419-0

Chiffi, D., & Pietarinen, A.-V. (2019b). The extended gabbay-woods schema and sientific practices. *Natural Arguments: A Tribute to John Woods on the Occasion of his 80th Birthday* (pp. 331–348). College Publications.

Chiffi, D., Pietarinen, A.-V., & Proover, M. (2020). Anticipation, abduction and the economy of research: The normative stance. *Futures, 115*, 102471. https://doi.org/10.1016/j.futures.2019.102471

Colombo, M., & Hartmann, S. (2015). Bayesian cognitive science, unification, and explanation. *The British Journal for the Philosophy of Science, 68*(2), axv036. https://doi.org/10.1093/bjps/axv036

Colombo, M., & Wright, C. (2016). Explanatory pluralism: An unrewarding prediction error for free energy theorists. *Brain and Cognition*. https://doi.org/10.1016/j.bandc.2016.02.003

Cristalli, C., & Pietarinen, A.-V. (2020). Abstraction and generalization in Peirce's logic of experimental science: Cases from the 19th-century scientific practice. *HOPOS: History of Philosophy of Science, 11*(1), 93–121. https://doi.org/10.1086/713087

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*(12), 1704–1711. https://doi.org/10.1038/nn1560

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation, 7*(5), 889–904.

Friston, K. J. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences, 13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

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. (2012a). The history of the future of the Bayesian brain. *NeuroImage, 62*(2), 1230–1233. https://doi.org/10.1016/j.neuroimage.2011.10.004

Friston, K. J. (2012b). A free energy principle for biological systems. *Entropy (Basel, Switzerland), 14*(11), 2100–2121. https://doi.org/10.3390/e14112100

Friston, K. J. (2013). Active inference and free energy. *Behavioral and Brain Sciences, 36*(03), 212–213. https://doi.org/10.1017/S0140525X12002142

Friston, K. J. (2018). Am I self-conscious? *Frontiers in Psychology, 9*, 579. https://doi.org/10.3389/FPSYG.2018.00579

Friston, K. J., & Frith, C. (2015). A duet for one. *Consciousness and Cognition, 36*, 390–405. https://doi.org/10.1016/J.CONCOG.2014.12.003

Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience, 6*(4), 187–214. https://doi.org/10.1080/17588928.2015.1020053

Friston, K. J., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience, SEP*. https://doi.org/10.3389/fnhum.2013.00598

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159*(3), 417–458. https://doi.org/10.1007/s11229-007-9237-y

Gabbay, D. M., & Woods, J. (2005). *The reach of abduction. Insight and trial*. Elsevier.

Gabbay, D. M., & Woods, J. (2006). Advice on abductive logic. *Logic Journal of the IGPL, 14*(2), 189–219.

Gaultier, B. (2017). The iconicity of thought and its moving pictures: Following the sinuosities of Peirce's path. *Transactions of the Charles S. Peirce Society, 53*(3), 374–399.

Giere, R. N. (1987). The cognitive study of science. In N. Nersessian (Ed.), *The process of science: Contemporary philosophical approaches to understanding scientific practice* (pp. 139–159). Kluwer Academic Publishers.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese, 172*(2), 269–281. https://doi.org/10.1007/s11229-009-9506-z

Hintikka, J. (2007). *Socratic epistemology: Knowledge: Explorations of knowledge-seeking through questioning*. Cambridge University Press.

Hohwy, J. (2014). The self-evidencing brain. *Noûs, 50*(2), 259–285. https://doi.org/10.1111/nous.12062

Jaynes, E. T. (2003/1995). *Probability theory: The logic of science*. Cambridge 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

Ladyman, J., & Ross, D. (2007). *Every thing must go*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199276196.001.0001

Magnani, L. (2009). *Abductive cognition: The epistemological and eco-cognitive dimensions of hypothetical reasoning*. Springer.

Magnani, L. (2017). *The abductive structure of scientific creativity*. Springer.

Ma, M., & Pietarinen, A.-V. (2018). Let us investigate! Dynamic conjecture-making as the formal logic of abduction. *Journal of Philosophical Logic, 47*(6), 913–945. https://doi.org/10.1007/s10992-017-9454-x

Maynard-Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature, 246*(5427), 15–18.

Mayo, D. G. (2005). Peircean induction and the error-correcting thesis. *Transactions of the Charles S Peirce Society*, *41*(2). https://doi.org/10.1353/csp.2011.0039

Park, W. (2017). *Abduction in context: The conjectural dynamics of scientific reasoning*. Springer.

Peirce, C. S. (1876). Note on the theory of the economy of research (R 1093). Report of the U. S. Coast Survey for 1876, pp. 197–201. Reprinted in CP 7.76–83.

Peirce, C. S. (1905). Letter draft to Lady Victoria Welby, July 16, 1905 (R L 493). The Peirce Papers, Harvard Houghton Library.

Peirce, C. S. (1913). An essay toward improving our reasoning in security and in uberty (R 682–4). The Peirce Papers, Harvard Houghton Library.

Peirce, C. S. (1931–1966). *The collected papers of Charles S. Peirce*. 8 vols., ed. by Hartshorne, C., Weiss, P., & Burks, A. W. Cambridge: Harvard University Press. Cited as CP followed by volume and paragraph number.

Peirce, C. S. (1967). Manuscripts in the Houghton Library of Harvard University, as identified by Richard Robin, *Annotated Catalogue of the Papers of Charles S. Peirce*, Amherst: University of Massachusetts Press, 1967, and in The Peirce Papers: A supplementary catalogue, *Transactions of the C. S.*

Peirce Society 7 (1971), 37–57. Cited as R or R L (letters) followed by manuscript number and, when available, page number.

Peirce, C. S. (1998). *The essential Peirce. Volume 2: 1893–1913*. Edited by The Peirce Edition Project. Indiana University Press. Cited as EP 2.

Pietarinen, A.-V. (2006). *Signs of logic: Peircean themes on language, games, and communication*. Springer.

Pietarinen, A.-V. (2014). The science to save us from philosophy of science. *Axiomathes, 25*, 149–166.

Pietarinen, A.-V., & Bellucci, F. (2014). New light on Peirce's conceptions of retroduction, deduction, and scientific reasoning. *International Studies in the Philosophy of Science, 28*(4), 353–373.

Pietarinen, A.-V. (2015). Recent studies on signs: Commentary and perspectives. *Sign Systems Studies, 43*(4), 616–650.

Pietarinen, A.-V. (2018). Conjectures and abductive reasoning in games. *Journal of Applied Logics - IfCoLog Journal of Logics and Its Applications, 5*(5), 1121–1144.

Pietarinen, A.-V. (2020). Abduction and diagrams. *Logic Journal of the IGPL, 29*(4), 447–468. https://doi.org/10.1093/jigpal/jzz034

Ramstead, M. J., Kirchhoff, M. D., & Friston, K. J. (2019). A tale of two densities: Active inference is enactive inference. *Adaptive Behavior*, 105971231986277. https://doi.org/10.1177/1059712319862774

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1), 79–87. https://doi.org/10.1038/4580

Rescher, N. (1976). Peirce and the economy of research. *Philosophy of Science, 43*(1), 71–98. https://doi.org/10.1086/288670

Rescher, N. (1978). *Peirce's philosophy of science: Critical studies in his theory of induction and scientific method*. University of Notre Dame Press.

Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience, 5*(2), 97–118. https://doi.org/10.1080/17588928.2013.877880

Shanahan, M. (2005). Perception as abduction: Turning sensor data into meaningful representation. *Cognitive Science, 29*(1), 103–134. https://doi.org/10.1207/s15516709cog2901_5

Weisberg, M. (2007). Who is a modeler? *The British Journal for the Philosophy of Science, 58*(2), 207–233. https://doi.org/10.1093/bjps/axm011

Wible, J. R. (1994). Charles Sanders Peirce's economy of research. *Journal of Economic Methodology, 1*(1), 135–160. https://doi.org/10.1080/13501789400000009

Wible, J. R. (2019). C. S. Peirce's theory of abductive expectations. *European Journal of the History of Economic Thought, 27*(1), 2–44. https://doi.org/10.1080/09672567.2019.1635179