

УДК: 165.62

DOI: 10.17223/1998863X/58/4

**I.F. Mikhailov**

## **INFERENCE AND REPRESENTATION: PHILOSOPHICAL AND COGNITIVE ISSUES**

*The paper is dedicated to particular cases of interaction and mutual impact of philosophy and cognitive science. Thus, philosophical preconditions in the middle of the 20th century shaped the newly born cognitive science as mainly based on conceptual and propositional representations and syntactical inference. Further developments towards neural networks and statistical representations did not change the prejudice much: many still believe that network models must be complemented with some extra tools that would account for proper human cognitive traits. I address some real implemented connectionist models that show how 'new associationism' of the neural network approach may not only surpass Humean limitations, but, as well, realistically explain abstraction, inference and prediction. Then I stay on Predictive Processing theories in a little more detail to demonstrate that sophisticated statistical tools applied to a biologically realist ontology may not only provide solutions to scientific problems or integrate different cognitive paradigms but propose some philosophical insights either. To conclude, I touch on a certain parallelism of Predictive Processing and philosophical inferentialism as presented by Robert Brandom.*

*Keywords: inference, representation, cognitive science, predictive processing, active inference, inferentialism, Brandom.*

### **1. The problems with distributed representations**

There is at least one important link between cognitive science and philosophy. The earlier attempts on science of mind, as it seems, were made in the times when philosophers mainly shared the 'propositional attitudes' jargon, and one of the primary cares of theirs was the question of whether contents of perceptions are conceptual and even propositional<sup>1</sup>. An answer thereto, which could be positive to any extent, would impose a view that linguistic structures are in some sense prerequisite to mind in general. This supposition shaped computational systems at the dawn of cognitive science, e.g., the widely known ACT family of software models designed by John Anderson. He himself described his initial conception like this:

In our evolution we may have developed or enhanced certain features to facilitate language, but once developed, these features were not confined to language and are now used in nonlinguistic activities. Thus the mind is a general pool of basic structures and processes, which has been added to under evolutionary pressure to facilitate language [1. P. 3].

However, further advancement of cognitive computational modeling brought about network architectures [2–4] that, in a couple of decades, developed into the impressive industry of artificial neural networks [5, 6], whose recognition competence and creative talents make for a kind of new epos nowadays. This development led to localization of the processing itself at some sub-linguistic and

---

<sup>1</sup> For the discussion see [36–38].

even sub-symbol levels of cognitive structures. The established conception of semantic ascending from elementary to complex signs was found endangered as networks did not identify particular cells to store and retrieve particular meanings. Lively discussions followed that will be considered below. Today it is more or less commonly agreed among cognitive scientists that the things familiarly addressed to as ‘representations’ are, in fact, a kind of statistical densities or distributions stored as vectors of values of multiple variables. This view is supported by the fact of parallel development of non-network statistical machines industry that is conventionally labeled ‘*Machine Learning*’ and is unambiguously preferred to that of neural networks by some central figures of computational cognitive science [7].

As far as I can tell, this new research landscape has not affected philosophical conventions yet, to any noticeable extent. Philosophers on both sides of (anti-)representationalist dividing line keep up with ‘conceptual’ and ‘propositional’ idioms spreading this bias over some part of cognitive students, too. Thus, some of practicing researchers take the baton from philosophers in believing that, to think like humans, artificial neural networks must be technically enhanced for them to be able to:

(1) build causal models of the world that support explanation and understanding, rather than merely solving pattern recognition problems; (2) ground learning in intuitive theories of physics and psychology to support and enrich the knowledge that is learned; and (3) harness compositionality and learning-to-learn to rapidly acquire and generalize knowledge to new tasks and situations [5. P. 1].

Such statements can be nothing but presupposed by the belief that, besides neuronal computations over statistical representations, there is still a specifically human way of information processing characterized by explaining, understanding, theorizing, producing new knowledge and generalizations, which requires, therefore, respective complementation of the plain network architecture to make for a ‘strong’, human-like AI. By analogy with the famous ‘folk psychology’, this discourse could be named ‘folk epistemology’, whose redundancy may be demonstrated by success of better and more consistent theories.

## **2. How can associations produce rational inferences?<sup>1</sup>**

The advancements of connectionism in the 1980s–1990s and, subsequently, those of modern artificial neural networks in biologically realistic explanations of cognitive capabilities posed new questions to both philosophers and empirical researchers. The need to reconsider some cornerstone concepts – namely, those of computations and representations – became apparent. As a reaction thereto, there was an attack, well-known in the history of science, by Jerry Fodor and Zenon Pylyshyn onto the semantic capabilities of connectionist networks [8]. The main objective of their onslaught was a supposed failure of connectionism to account for systematicity, compositionality and productivity of human cognitive resources including, but not restrained to, language. The authors put forward arguments based on the alleged combinatorial nature of mental representations, in the light of which productivity, systematicity and compositionality presumably inherent in the human language were extrapolated to the entire cognitive sphere. In their opinion,

---

<sup>1</sup> I owe the idea of this section to Anna Khromchenko (Tomsk State University) who involved me in a fruitful discussion on those matters in the cold rainy summer of 2019.

this argument shows that the architecture of consciousness is not connectionist at the cognitive level, since connectionist representations do not demonstrate these properties. But, for my present purpose, their overall dismissal of connectionism as a new associationism is a point of lasting philosophical importance despite their paper being aged.

Fodor and Pylyshyn contrapose classical, Turing / von Neumann-based computational models and those of connectionists as, respectively, structure-sensitive and frequency-sensitive ones, so that connectionist networks may master logical inferences only because of a certain frequency of joint occurrence of ideas and not their formal (structural) properties. This, in their view, is a step back to good old Associationism with some newer technical amendments, which are of no help in increasing its poor explanatory power. Here is how they put it:

Associative strength was not, however, presumed to be sensitive to features of the content or the structure of representations per se. Similarly, in Connectionist models, the selection of an output corresponding to a given input is a function of properties of the paths that connect them (including the weights, the states of intermediate units, etc). And the weights, in turn, are a function of the statistical properties of events in the environment (or of relations between patterns of events in the environment and implicit ‘predictions’ made by the network, etc.) But the syntactic / semantic structure of the representation of an input is not presumed to be a factor in determining the selection of a corresponding output since, as we have seen, syntactic / semantic structure is not defined for the sorts of representations that Connectionist models acknowledge [8. P. 20–21].

In a nutshell, here is how the situation looked then and still does. You have got some exclusive human cognitive powers to explain, specifically, the fact that they are structured, rule-based, combinatorial and productive. You try some old-fashioned associationism, and it fails. You look at some theory that pretends to be state-of-the art, and it reminds you the same old-fashioned associationism, so you expect it to fail either. Then you stay with your preferred theory that simply postulates structuredness, rule-compliance, systematicity, compositionality and productivity of the things you try to scientifically explain. And you consider it to be the real solution.

A lot of cognitive scientists and philosophers have replied to Fodor and Pylyshyn’s challenge since it was published<sup>1</sup>. But we had better look at some actually implemented connectionist models that overtly addressed the well-known limitations of Hume’s associationism. For reasons that remain unclear to me, these publications have gone mainly unnoticed so far, except for a couple of citations and quick mentions in encyclopedias.

Mark Collier [9] highlights two Hume’s hypotheses aimed to explain the paradox of ‘continued existence’, that is our inferred belief in the consistent identity of an object that produces inconsistent series of impressions: the one he refers to as ‘conflation account’ that relies on qualitative resemblances of interrupted series, and the other labeled ‘assimilation account’ that puts forward the

---

<sup>1</sup> Paul Smolensky [39] argued that the critics did not take into account the distributed nature of connectionist representations and that connectionist studies should offer new formalizations of fundamental computational concepts. His argument was subsequently countered by [40]. William Ramsey [41] showed that unidirectional networks with backward propagation of error do not need the concept of representation as such. See also [42–45] and many others.

constructive role of imagination. The first one he dismisses as mainly unsustainable, while the other gets proper consideration. Collier believes that “not only does connectionist theory allow us to complete the assimilation hypothesis by supplying the missing principles of the imagination, but connectionist methodology provides the experimental conditions under which the hypothesis can be implemented and tested” [9. P. 162]. He proves it by an experiment with a real simple recurrent network (SRN) that by design is able to extrapolate previously experienced patterns on current unsteady phenomena and, therefore, to represent an assumed underlying object as consistently existing. And, as he points out, experimentation with computer models is advantageous in the sense that, contrary to that with living subjects, we are able to look inside the ‘brain’ and find out that the anticipating representation of the reoccurring object, as it were, thickens and matures with newer learning cycles. As he concludes, ‘experimental results demonstrate that the belief in continued existence can arise solely from the interaction of sensory information with the principles of an information-processing mechanism’ [9. P. 164].

Dan Ryder and Oleg V. Favorov [10] go even deeper in searching for neural computational grounds for associationist explication of rational capabilities. First of all, they briefly engage in a philosophical dispute of whether a brain performs some unified task, or it is a ‘hodgepodge’ of different modules whose multidirectional activities give rise to emergent cognitive features. As for me, I would rather stay with the latter option, but the authors choose the former one, and the single mission of the brain (or, at least, of its cortex) is supposed to be prediction<sup>1</sup>. Their working hypothesis states that the dendritic trees of individual pyramidal cells in the cerebral cortex are structured in such a way that they are in a position to learn ever deeper regularities in the environment represented in multiple variables, not being restrained to associated pairs of ideas considered in Hume’s analysis. As for the latter, the authors believe that his pairwise associationism inherited or mirrored Newtonian mechanicism in that it aspired to find basics of the rational in some simple irrational devices. The problem is that, to derive reason from impressions, Hume had to postulate the ability for abstraction, for which, unlike his model of associations, he had no mechanistic explanation. That is why, long since, it was “not Hume’s mechanistic theory, but Turing’s, that has had some success in modeling reason. But the brain rarely plays more than a small part (if any) in such theories, and they do not sit well with empiricism” [10. P. 163].

As Ryder and Favorov see it, connectionist modeling may compensate this shortage. Their ambitions go as far as to show that

In fact, the mechanism of association and the mechanism that replaces abstraction turn out to be identical, which results in a unified explanation of two fundamental mental processes: rational transitions in thought (reasoning) and representation acquisition. This yields the beginnings of a neural theory, not only of the brain, but also of the mind [10. P. 164].

Ryder and Favorov’s SINBAD model<sup>2</sup> involves substitution of dendrites with what is seen as their functional approximations, namely, two backpropagation neural networks interrelated via a third one substituting a cell’s soma. Each of the

---

<sup>1</sup> Please note that they stated this a decade before the now acclaimed Predictive Processing became the order of the day.

<sup>2</sup> ‘SINBAD’ stands for ‘a Set of INteracting BAcKpropagating Dendrites’.

dendrites has two input channels that are tuned to ‘perceive’ the presence of some object mutually exclusively (by a *XOR*-function). Their outputs are put together by a certain set of functions to produce the cell’s output. The latter is looped back to alter each of the dendrite’s input. The error backpropagating signal represents the difference of the dendrite’s output and that of the cell.

The process of adjusting the weights of connections inside the two networks will continue until each dendrite learns to predict, basing on its own input data, the responses of the other dendrite to its input data. By teaching each other, dendrites of the cell can tune to different, but correlated functions, capable of revealing the order of the environment. The cell as a whole can learn to recognize the source of these correlated functions as an ordered property of its sensory environment. The first ability implements an association of functions according to conditions, in contrast to a simple Hume association. The second ability (which is actually just the flip side of the first) implements the process of obtaining representation, which replaces Hume’s abstraction.

Activities that begin on the periphery as sensory data are spread across the network in a way that implements not only simple induction, but also deductive reasoning and an inference to the best explanation. This shows how a simple biologically realistic mechanism can reveal the complexity of order in nature and use this knowledge for the vital task of prediction [10. P. 191].

Looking ahead at ideas considered below, it must be noted that the SINBAD theory may offer a mechanistic explanation of a possible neurodynamic implementation of Karl Friston’s ‘generative models’.

### **3. Predictive processing: generative models and sensory updates**

#### **3.1. Core idea and predecessors**

‘Predictive Processing’ (PP) is, as many believe, one of the most influential and explanatory-powerful cognitive approach nowadays [11–13].

Eventually shaped in the mid-2010s, the new paradigm claims to overcome the limitations of previous approaches. One of its founders, the British neuroscientist Karl Friston provides a biologically plausible explanation based on a conception of updating internal representations using sensory samples. The theory postulates the existence of generative models that produce downward predictions, which are met and compared with upward representations at a lower level in order to calculate the prediction error [14. P. 392]. The natural urge to minimize the difference of the predictive representation and the incoming data makes up the gist of the so called ‘free energy principle’ (FEP), more on which below.

The PP proponents derive the main principles of their approach from long-term philosophical and psychological doctrines, such as those of Alhazen, Kant and Helmholtz [15. P. 210]. As for the latter, it goes back to his idea of ‘unconscious inferences’. Those are formed in the early life experience and constitute the basis of many perceptual phenomena. According to Helmholtz, we tune our senses to distinguish things that affect them with maximum accuracy. Perception, thus, is a result of a meeting of external input with what the individual has already learnt [16]. The physicist accounts also for the very notion of ‘free energy’ [17]. In the late 20th century, his name was given to the so-called Helmholtz Machine – a hierarchical unsupervised learning algorithm that is capable of identifying structures underlying various data patterns [18].

### 3.2. The main concepts: thermodynamics, Bayesian statistics and information theory

PP comprises a sophisticated web of interrelated notions and concepts, which is not an easy way to break through. As far as I can tell, it builds upon a very basic idea of ‘predictive coding’, to which it adds concepts borrowed from thermodynamics, statistics and information theory. ‘Predictive coding’ amounts to positing generated a priori models that get successively updated by weighing prediction errors, the latter being difference of prediction and newly acquired data. Usually, when someone nominates various past theories as predecessors of PP, such as, e.g., Piotr Anokhin’s ‘functional systems’ [19], they address this very basic approach, but not mathematical and other subtleties comprised in the full-fledged PP, which are very important for determining the explanatory and predictive scope of the theory.

What PP adds thereto is, on the one hand, the conception of *precision optimization* that modulates prediction errors computation at different levels of the system. Precision of samples is optimized by previously learnt experience, thus demanding the statistical framework called ‘empirical Bayes’ for calculations. This implies that the processing is multi-level and context-dependent and, thereby, the processing system, be it a cell, a human or a robot, is capable of approximating hierarchical empirical Bayes inference, owing to which it functions better in adapting to fuzzy and everchanging environment than a system executing only exact Bayes inference [15. P. 213].

On the other hand, PP is extensively based on the concept of *active inference*. To minimize prediction errors, a system may, for one, update predictive models to comply with sensory input. But, alternatively, it may be active in sampling the environment in search for data that fit the prediction better. This means just action, hence the term. According to Friston, a living organism may be easier explained as an acting system, if we suppose that the triggers for active inference are proprioceptive data, because those may be directly functionally linked to reflex arcs [20–22]. Thus, PP is claimed to be a unified theoretical framework capable of explaining both perception and action.

The above-mentioned conceptions are process theories that appeal to real or modeled mechanisms and may be, therefore, directly falsified. But at the heart of PP lies a general principle that functions like the most known principles of natural science: it shapes explanations of process theories but is not directly falsifiable itself. It is the so-called *free energy principle*, often referred to as FEP. It postulates ‘minimizing the free energy’ as the principal urge of any self-sustaining or autopoietic systems, living systems being the main exemplification thereof. Borrowed from thermodynamics, the concept of free energy is adopted and explained by Karl Friston as the ‘surprise’ to be reduced [23]. Large amount of surprise, i.e., mismatch of a generated predictive model and sensory data, is too costly for a cognitive, or broadly – a living – system and needs to be reduced to the minimum available. This need triggers both perceptual and active inference.

Equipped with its full toolbox, PP leads not only to interesting empirical explanations, like those of mood change or schizophrenia, but also to some philosophical implications. Thus, the idea of perception as constant inference from sensory data to their probable causes delivers an interesting re-formulation of such philosophers’ favorite subjects as body image, sense of ownership, and bodily self-

awareness. The said phenomena are naturally explained as the inferred causes of interoceptive and proprioceptive sensations. According to Jacob Hohwy, the very Self in this context may be presented “as a subset of the inferred causes of sensory input that relates to own actions, and, consequently, a possibility for discussing whether such a set of causes is deserving of the label ‘self’ ” [15. P. 217]. So much extended PP obviously subsumes familiar subjects of Merleau-Ponty-style phenomenology that was famously the methodological ground for enactive and embodied cognitive science [24]. Interestingly, such a prominent proponent of introducing phenomenology into cognitive science as Thomas Metzinger has become one of PP-enthusiasts and is now one of the editors of the comprehensive thematic web resource [25] dedicated to PP and containing a kind of online encyclopedia on the topic.

### 3.3. Generative models and communication

Another conceptual pillar of PP is positing a set of *generative models (GM)* in the brain that are responsible for producing what may be taken as possible representations of the environment. The notion of such models involves multi-level organization, such that

the hierarchical structure allows priors at one level to be supplied by posteriors at a higher level. Sensory data are assumed to reside only at the lowest level in the hierarchy, and the highest level is assumed to generate only spontaneous random fluctuations [17. P. 75].

The same authors heavily relying on complex mathematical calculi provide an informal definition of GM as ‘a description of causal dependencies in the environment and their relation to sensory signals’ [17. P. 61]. In a more detailed manner, GMs are explicated in terms of recognition density (*R-density*) and generative density (*G-density*), meaning probability densities (distributions) in both cases. An organism is supposed to model likelihood of environmental variables, which is expressed as R-density. But, to do so, it must be capable of evaluating general dependencies of incoming sensory data from environmental states. A statistical model of those dependencies is expressed as G-density. Then interdependency of both kinds of densities makes for a generative model. As for a mechanistic implementation of these computational models, they are “instantiated, and parameterised, by physical variables in the organism’s brain such as neuronal activity and synaptic strengths, respectively” [17. P. 57]. As one may conclude thereby, PP actually builds on, amends and enriches the initial connectionist doctrine, utilizing, in particular, other formal tools while basing on the same ontology.

Friston demonstrates an interesting application of this theoretical framework to modelling language communication, which previously was the realm of classical symbolist cognitive science. According to him, the criteria for evaluating and fine-tuning the interpretation of another’s behavior are the same that underlie actions and perceptions in general, namely, minimization of prediction errors. The concept of communication in PP is based on a generative model, or narrative, which is shared by agents exchanging sensory signals.

As Friston puts it, models based on hierarchical attractors that generate various categories of sequences allow closing the hermeneutic circle by simply updating generative models and their predictions in order to minimize prediction

errors. It is important to note that these errors can be calculated without even knowing the true state of another, which thereby solves the problem of hermeneutics [26. P. 129–130].

Friston and colleagues built a computer emulation of two songbirds using software agents whose tweets were generated by some attractor-based models and recursively refined in the process of mutual listening. The model showed that birds follow the narrative produced by dynamic attractors in their generative models, which were synchronized through sensory exchange. This means that both birds can sing ‘from one music sheet’ while maintaining a consistent and hierarchical structure in their overall narration. It is this phenomenon that Friston associates with communication [14. P. 400]. Generative models used to determine one’s own behavior can be used to derive the beliefs and intentions of the other, provided that both sides have fairly similar generative models. This perspective creates representations of a set of intentional acts and narratives, suggesting a collective narrative shared by communicating agents [14. P. 401].

One may suggest that the explanatory capabilities of PP span not only over issues of psychology and conventional philosophy of mind, but over the newer subject-matters of ‘social mind’ and social cognition as well. Certainly, such a universality may raise concerns of falsifiability of the theory. For a brief take on that matter see [15. P. 221]. But, generally, such a worry may relate to the question, if each and every behavior may be presented as guided by the subject’s statistical predictions. It really seems that a straightforward refutation is hard to imagine in this case, but there is always a place for a better theory to demonstrate its greater explanatory potential. Furthermore, PP is still too young as a theory: while performing to a greater or lesser success in actual experiments it will inevitably be met with an urge to give detailed mechanistic accounts of all the statistic models implemented. Those accounts will certainly be essentially falsifiable.

#### **4. Philosophical Inferentialism: top-down semantics and anti-empiricist pragmatism**

As we already know, in order to advance the free energy minimization, a cognitive agent actively engages in the intercourse with its environment. This part of the system’s functioning is labeled ‘active inference’ [21, 22, 27]. But not only this terminal part is inferential – the whole system functions by gradual inferring consequential options (‘priors’) for every next lower level of the cognitive machine. That is why it would be interesting to compare this cognitive inferential view with *inferentialism* as an influential philosophical school of recent to see their points of intersection, if any.

One of the most known proponents on the philosophical inferentialism is Robert Brandom [28, 29]. Inspired by Frege, Sellars and, partly, by later Wittgenstein, Brandom strongly opposes any representationalist accounts of mind and language.

Representationalism, in his view, inherits to the classical empiricism of the New Age European philosophy. It imposes a kind of bottom-up semantics by stating that the meaning of a sentence is a function of its sub-sentential constituents. Brandom counters representationalism with a view, according to which the meaning of a sentence boils down to its inferential roles in various parts of the discourse it is engaged in. And, correspondingly, meanings of its constituent



terms are derived from the meaning of the whole. This top-down semantic approach is in line with classical European rationalism and even German idealism exemplified by Hegel, whom Brandom admires a lot.

Interestingly, in 2009 Brandom gave a talk entitled '*How Analytic Philosophy Has Failed Cognitive Science*' [30]. There he claimed that analytic philosophy could have but did not explain to cognitive scientists the importance of *the conceptual* in the proper sense. According to him, Aristotelian logic was based on classification as the only – and very poor – model of rational inference. Founders of cognitive science, being unfamiliar with the current achievements of analytic philosophy, missed the gist of Fregean revolution and based the newly born science on classification as the principal cognitive mechanism. Modern philosophers could have helped them in tying analytically found grades of conceptual advancement with actual developmental stages of human and animal psyche. But they did not, at least to the date of his talk.

Like it or not, Brandom himself could have indicated a way out of this failure. The important part of his rich doctrine is constituted of what he refers to as *pragmatism* [31, 32]. Though stemming from different premises than rationalism<sup>1</sup>, classical American pragmatism opposed the straightforward picture-like conception of experience. According to Pierce, James and Dewey, an organism is not a purified receiver of impressions but a defining part thereof<sup>2</sup>. What and which way we may experience is determined by what we are.

Could anti-representationalism be the meeting point for PP and inferentialism? PP is broadly considered an ultra-representationalist cognitive theory [33]. But there are some grounded claims that there are at least two versions of the doctrine: the representational one and the one preliminarily denoted 'enactive', stressing 'active inference' narrative [34]. Brandom could have inferred that his inferentialist-pragmatist philosophy may be a saving bridge between the two most influential cognitive paradigms of today: Predictive Processing and what many refer to as *4E-Cognition*, meaning 'embodied, embedded, enactive, and extended'. To this end, one could dismiss the purely representationalist account of the PP-doctrine and turn to its versions compatible with enactive and embodied cognition [34] by exploiting the outlined conceptual chain: *active inference – inferentialism – pragmatism – enactivism*. But this is a vast underexplored field so far, quite out of the scope of the present paper. And it is too early to take positions here, in my opinion.

But there is, nevertheless, at least one interesting study that may properly dispose representational idioms inside PP discourse. In [35] the author uses PP to take on the so called Sellarsian dilemma that unveils justification issues with regarding sensory states as epistemically basic. He shows how sensory signals may conditionally justify perceptual predictions and, at the same time, play an unconditionally justificational role within perceptual learning. That is, sensory signals may play a crucial role in justifying representational states while being non-representational themselves.

---

<sup>1</sup> I would propose that, unlike modern-day inferentialism, classical European rationalism differs from empiricism not in an anti-representationalist stance, but only in a different understanding of the source of representations. Therefore, Brandom's heirloom of Hegel may be regarded as only partial.

<sup>2</sup> More on this in [34].

All in all, there are conceptual relations linking together philosophical inferentialism and PP as soon as the latter reaches issues of language and the properly logical inference. The knowledge of cognitive implementations of the said conceptual schemas may not be a crucial proof thereof, but undoubtedly a noticeable support: it is hard to uphold any philosophical alternatives knowing actual mechanisms.

## 5. Conclusion

I have attempted to review a couple of intertwined discussions going on in philosophy of cognitive science these days. As I said, it may be too early to bring about any final judgements, but some tendencies seem quite obvious.

First of all, the former gap between explanatory models of psychology and neuroscience is being bridged. The latest theoretical trends do care about biological realism and mechanicism of their explanations. Second, interdisciplinary integration is manifest in the latest research: cognitive scientists borrow conceptions and approaches from the likes of biology and thermodynamics, while applying all the more complicated formal tools, such as statistics, probabilistic logic and information theory.

And, at last, what refers directly to the problem under question here is the change in methodological or metatheoretical thinking that makes former disputes, such as classicism vs. connectionism or inferentialism vs. representationalism, outdated and ungrounded. When logomachy is replaced with concrete mechanistic and mathematically backed models, one may find proper places for both inference and representation, as soon as they are operationally defined.

Cognitive studies of the kind that is not predefined by philosophically imposed presuppositions and prejudices but proceeds from the urge to look for all possible implementations of a function under question, may, in their turn, teach an important lesson to philosophers. Problems that seem black-and-white or generally unsolvable for a traditional rationalist *A-or-~A* view may occur to be on the way to solution under some quantitative or probabilistic approaches. If we do not know how to deduce, say, inferential capabilities from associative principles, that doesn't mean that the nature shares our ignorance. Unlike us, it had billions of years of trials and errors, of mutability and heredity, of all that we call evolution, which is nothing but one enormous algorithmic machine of statistic inference, incidentally invented and multiplied being embedded in myriads of cells, brains and societies.

## References

1. Anderson, R. (1983) *The Architecture of Cognition*. Hillsdale, NJ, US: Harvard University Press.
2. Pinker, S. & Prince, A. (1988) On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*. 28(1–2). pp. 73–193. DOI: 10.1016/0010-0277(88)90032-7
3. Rumelhart, D.E. & McClelland, J.L. (1986a) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1. Cambridge, MA, USA: MIT Press.
4. McClelland, J.L. & Rumelhart, D.E. (1986b) *Parallel Distributed Processing: Explorations in the Microstructure*. Vol. 2. Cambridge, MA, USA: MIT Press.
5. Lake, B.M., Ullman, T.D., Tenenbaum, J.B. & Gershman, S.J. (2017) Building machines that learn and think like people. *Behavioral and Brain Sciences*. 40. p. e253. DOI: 10.1017/S0140525X16001837
6. Mayor, J., Gomez, P., Chang, F. & Lupyan, G. (2014) Connectionism coming of age: legacy and future challenges. *Frontiers in Psychology*. 5. pp187. DOI: 10.3389/fpsyg.2014.00187

7. Poggio, T. (2012) The Levels of Understanding Framework, Revised. *Perception*. 41(9). pp. 1017–1023. DOI: 10.1068/p7299
8. Fodor, J.A. & Pylyshyn, Z.W. (1988) Connectionism and cognitive architecture: a critical analysis. *Cognition*. 28(1–2). pp. 3–71. DOI: 10.1016/0010-0277(88)90031-5
9. Collier, M. (1998) Filling the gaps: Hume and connectionism on the continued existence of unperceived objects. *Hume Studies*. 25(1–2). pp. 155–170.
10. Ryder, D. & Favorov, O.V. (2001) The New Associationism: A Neural Explanation for the Predictive Powers of Cerebral Cortex. *Brain Mind*. 2(2). pp. 161–194. DOI: 10.1023/A:1012296506279
11. Clark, A. (2015) Radical predictive processing. *The Southern Journal of Philosophy*. 53(S1). pp. 3–27. DOI: 10.1111/sjp.12120
12. Friston, K. (2012) Prediction, perception and agency. *International Journal of Psychophysiology*. 83(2). pp. 248–252. DOI: 10.1016/j.ijpsycho.2011.11.014
13. Hohwy, J. (2014) *The Predictive Mind*. Oxford University Press.
14. Friston, K. & Frith, C. (2015) A Duet for one. *Consciousness and Cognition*. 36. pp. 390–405. DOI: 10.1016/j.concog.2014.12.003
15. Hohwy, J. (2020) New directions in predictive processing. *Mind Language*. 35(2). pp. 209–223. DOI: 10.1111/mila.12281
16. Helmholtz, H. von (2013) *Treatise on Physiological Optics*. Dover Publications.
17. Buckley, C.L., Kim, C.S., McGregor, S. & Seth, A.K. (2017) The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*. 81. pp. 55–79. DOI: 10.1016/j.jmp.2017.09.004
18. Dayan, P., Hinton, G.E., Neal, R.M. & Zemel, R.S. (1995) The Helmholtz Machine. *Neural Computation*. 7(5). pp. 889–904. DOI: 10.1162/neco.1995.7.5.889
19. Egiazaryan, G.G. & Sudakov, K.V. (2007) Theory of functional systems in the scientific school of P.K. Anokhin. *Journal of the History of the Neurosciences*. 16(1–2). pp. 194–205.
20. Friston, K.J., Daunizeau, J., Kilner, J. & Kiebel, S.J. (2010) Action and behavior: A free-energy formulation. *Biological Cybernetics*. 102(3). pp. 227–260. DOI: 10.1007/s00422-010-0364-z
21. Friston, K., Mattout, J. & Kilner, J. (2011) Action understanding and active inference. *Biological Cybernetics*. 104(1–2). pp. 137–160. DOI: 10.1007/s00422-011-0424-z
22. Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O’Doherty, J. & Pezzulo, G. (2016) Active inference and learning. *Neuroscience and Biobehavioral Reviews*. 68. pp. 862–879. DOI: 10.1016/j.neubiorev.2016.06.022
23. Friston, K. (2009) The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*. 13(7). pp. 293–301. DOI: 10.1016/j.tics.2009.04.005
24. Varela, F.J., Rosch, E. & Thompson, E. (1992) *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.
25. Metzinger, T. & Wiese, W. (n.d.) *Philosophy & Predictive Processing*. [Online] Available from: <https://predictive-mind.net> (Accessed: 30th May 2020).
26. Friston, K.J. & Frith, C.D. (2015) Active inference, communication and hermeneutics. *Cortex*. 68. pp. 129–143. DOI: 10.1016/j.cortex.2015.03.025
27. Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T. & Dolan, R.J. (2013) The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience*. 7(SEP). p. 598. DOI: 10.3389/fnhum.2013.00598
28. Brandom, R.B. (2001a) *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
29. Brandom, R.B. (2001b) *Articulating Reasons: an Introduction to Inferentialism*. Harvard University Press.
30. Brandom, R.B. (2009) How analytic philosophy has failed cognitive science. *CEUR Workshop Proceedings*. 444.
31. Brandom, R.B. (2008) *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford University Press.
32. Brandom, R.B. (2011) *Perspectives on Pragmatism: Classical, Recent, and Contemporary*. Harvard University Press.
33. Clowes, R.W., Gärtner, K. & Clowes, R.W. (2017) Enactivism, Radical Enactivism and Predictive Processing: What is Radical in Cognitive Science? *Kairos Journal of Philosophy and Science*. 18(1). pp. 54–83. DOI: <https://doi.org/10.1515/kjps-2017-0003>
34. Williams, D. (2018) Pragmatism and the predictive mind. *Phenomenology and the Cognitive Science*. 17(5). pp. 835–859. DOI: 10.1007/s11097-017-9556-5

35. Gładziejewski, P. (2017) The Evidence of the Senses – A predictive processing-based take on the Sellarsian dilemma. In: Metzinger, T. & Wiese, W. (eds) *Philosophy and Predictive Processing*. Vol. 15. Frankfurt am Main: MIND Group. DOI: 10.15502/9783958573161
36. Crane, T. (2009) Is Perception a Propositional Attitude? *The Philosophical Quarterly*. 59(236). pp. 452–469. DOI: 10.1111/j.1467-9213.2008.608.x
37. Runzo, J. (1993) The Propositional Structure of Perception. In: Runzo, J. (ed.) *World Views and Perceiving God*. London: Palgrave Macmillan. pp. 3–22.
38. Byrne, A. (2009) Experience and Content. *The Philosophical Quarterly*. 59(236). pp. 429–451. DOI: 10.1111/j.1467-9213.2009.614.x
39. Smolensky, P. (1988) On the proper treatment of connectionism. *Behavioural and Brain Sciences*. 11(1). pp. 1–23. DOI: 10.1017/S0140525X00052432
40. Fodor, J. & McLaughlin, B.P. (1990) Connectionism and the problem of systematicity: why Smolensky's solution doesn't work. *Cognition*. 35(2). pp. 183–204. DOI: 10.1016/0010-0277(90)90014-B
41. Ramsey, W. (1997) Do Connectionist Representations Earn Their Explanatory Keep? *Mind & Language*. 12(1). pp. 34–66. DOI: 10.1111/j.1468-0017.1997.tb00061.x
42. Matthews, R.J. (1997) Can Connectionists Explain Systematicity? *Mind & Language*. 12(2). pp. 154–177. DOI: 10.1111/j.1468-0017.1997.tb00067.x
43. Gomila, A., Travieso, D. & Lobo, L. (2012) Wherein is Human Cognition Systematic? *Minds and Machines*. 22(2). pp. 101–115. DOI: 10.1007/s11023-012-9277-z
44. Phillips, S. & Wilson, W.H. (2010) Categorical compositionality: a category theory explanation for the systematicity of human cognition. *PLoS Comput Biol*. 6(7). p. e1000858. DOI: 10.1371/journal.pcbi.1000858
45. Aizawa, K. (1997) Explaining Systematicity. *Mind & Language*. 12(2). pp. 115–136. DOI: 10.1111/j.1468-0017.1997.tb00065.x

**Igor F. Mikhailov**, Institute of Philosophy, Russian Academy of Sciences (Moscow, Russian Federation).

E-mail: ifmikhailov@gmail.com

*Vestnik Tomskogo gosudarstvennogo universiteta. Filosofiya. Sotsiologiya. Politologiya – Tomsk State University Journal of Philosophy, Sociology and Political Science*. 2020. 58. pp. 34–46.

DOI: 10.17223/1998863X/58/4

## INFERENCE AND REPRESENTATION: PHILOSOPHICAL AND COGNITIVE ISSUES

**Keywords:** inference; representation; cognitive science; predictive processing; active inference; inferentialism; Brandom.

The article attempts to identify some significant and parallel trends happening in both philosophy and cognitive science during recent decades. The author's conjecture is that, in spite of crucial turns occurred in cognitive science with the advance of neural networks and statistical explanations, some outdated philosophical commitments and presuppositions prevail over the facts and technologies making some researchers stay with their strong belief that this 'new associationism' and frequency-based drift is not sufficient to explain human higher cognitive capacities. According to some of them, network or statistical AI models must be amended with conceptual or inferential tools in order to match the human intelligence. In a series of reviews, the author presents evidences borrowed from empirical and conceptual studies revealing principal continuity of perception and conceptual inference, as well as the underlying mechanisms of their gapless connection. The reason for the opposite views and presuppositions lingering in the cognitive and philosophical literature is, according to the author, sticking to the inherited logically determined methods, while the reality under investigation is principally stochastic. That is, good old conceptual analysis is getting less and less useful in coping with cognitive matters. Jerry Fodor and Zenon Pylyshyn were the first in the onslaught on the newly born connectionism in the 1980s trying to show that, as its explanations were frequency-based but not sensitive to formal properties, it is no more than a rebirth of long-forgotten Humean associationism. The latter, in their view, is incapable of deriving conceptual and inferential capacities from repetition-based associations of ideas. The author counterposes two connectionist studies published in the early 2000s exemplifying neural network models capable of abstraction and prediction without resorting to symbolic or conceptual or any other linguistically induced tools. Then the author proceeds to the more recent achievement known as 'predictive processing'. This paradigm abandons the perception/inference dualism by positing that the main principle governing cognitive and, broader, biological processes is 'free energy'

minimization. ‘Free energy’ here is a metaphor standing for the difference of anticipated and actual data that an organism happens to acquire and process. Anticipation, or prediction, is facilitated by its internal attractor-based generative models that are capable of updating once ‘free energy’ exceeds a certain threshold. Therefore, having an excessive amount of ‘free energy’, an organism can either update its generative models, or resort to the so-called ‘active inference’, which stands for just action, in order to match prediction with input. Lastly, the author compares this predictive inferential approach to philosophical inferentialism advocated by Robert Brandom. Anti-representationalist stance of the latter resonates with some PP-proponents who state that sensory signals may conditionally justify perceptual predictions and, at the same time, play an unconditionally justificational role within perceptual learning. That is, sensory signals may play a crucial role in justifying representational states while being non-representational themselves. In this way, the classical European rational/empirical distinction loses its gist, at least at the level of an individual.