

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Brain Functional Architecture and Human Understanding

Yan M. Yufik

“Reagan. What need one?”

King Lear. O, reason not the need: our basest beggars

Are in the poorest thing superfluous:

Allow not nature more than nature needs,

Man's life's as cheap as beast's...”

William Shakespeare. King Lear, Act 1, Scene 4

Abstract

The opening line in Aristotle's *Metaphysics* asserts that “humans desire to understand”, establishing understanding as the defining characteristic of the human mind and human species. What is understanding and what role does it play in cognition, what advantages does it confer, what brain mechanisms are involved? The Webster's Dictionary defines understanding as “apprehending general relations in a multitude of particulars.” A proposal discussed in this chapter defines understanding as a form of active inference in self-adaptive systems seeking to expand their inference domains while minimizing metabolic costs incurred in the expansions. Under the same proposal, understanding is viewed as an advanced adaptive mechanism involving self-directed construction of mental models establishing relations between domain entities. Understanding complements learning and serves to overcome the inertia of learned behavior when conditions are unfamiliar or deviate from those experienced in the past. While learning is common across all animals, understanding is unique to the human species. This chapter will unpack these notions, focusing on different facets of understanding. The proposal formulates hypotheses regarding the underlying neuronal mechanisms, attempting to assess their plausibility and reconcile them with the recent ideas and findings concerning brain functional architecture.

Keywords: neuronal mechanisms, consciousness, understanding, brain function, functional architecture, neuronal correlations of understanding

1. Introduction

The concept of ‘mental models’, i.e. memory constructs acting as “small-scale models of reality” intervening between stimuli and responses was introduced in [1], and subsequently elaborated by multiple authors applying the concept in the context of various disciplines [2–6]. More general, domain-invariant theories conceptualize models as inferential frameworks enabling deductive and other forms of reasoning [7, 8], in particular, reasoning by analogy [9].

The theory of understanding discussed in this chapter (the VAN theory formulated in [10–12]) centers on the notions of self-adaptive processes in virtual associative networks (VAN) and defines understanding as a human-specific form of active inference subsumed under the principles of active inference and variational free energy minimization advanced in [13, 14]. The theory contends that curbing metabolic costs and regulating the dynamics of energy processes in the brain have been critical factors in the evolution of intelligence, culminating in the emergence of mental modeling mechanisms in humans that made possible explosive growth in the variety of activities a person can engage in without exploding either the number of neurons and/or the metabolic costs of neuronal processes necessary for organizing those activities. According to the theory, mental models are simultaneous memory structures imposing tight constraints on their constituent components and thus sharply reducing the number of degrees of freedom available to them. Reduction in the number of degrees of freedom minimizes the amount of processing in performing cognitive tasks, yielding two interrelated benefits: curbing energy demands and giving rise to abilities that define human intelligence and are inherent in the understanding capacity, i.e. prediction, explanation, and planning.

These ideas are explored in the present chapter, heeding the advice attributed to Einstein and suggesting that, when pondering a problem, the bulk of the effort needs to be spent on formulating the problem (as clearly as possible). Due to a confluence of circumstances, cognitive science has been downplaying the role of understanding in cognitive performance. The main thrust in this chapter is to examine and elevate that role. The chapter is organized in four parts. Section 2 reviews challenges to understanding posed by different tasks, Section 3 starts with an excursion into evolutionary history, focusing on differences in cognitive performance making human intelligence discontinuous with that of the other species, and Section 4 outlines a theory of understanding, building on the notions introduced in the preceding parts. Section 5 presents a discussion and brief concluding remarks.

2. Anatomy of understanding

Understanding involves grasping relations between entities, which boils down to fitting representations of these entities into simultaneous memory structures (mental models) that sharply reduce the number of degrees of freedom available to them. Illustrating how these processes operate in the understanding of literary works will help clarify the ideas.

2.1 Understanding Shakespeare

The corpus of literary work by William Shakespeare includes 37 plays and over 150 sonnets and poems. It has been estimated that a legion of monkeys with as many members as there are protons in the observable universe, each monkey having a typewriter and hitting randomly at the keys, would need the amount of time more than three hundred and sixty thousand orders of magnitude longer than the age of the universe in order to have a negligibly small chance (1 in 10^{500}) of having typed a single play (https://en.wikipedia.org/wiki/Infinite_monkey_theorem).

The adult human brain comprises 86 billion neurons and 85 billion non-neuronal cells [15] which are vanishingly small numbers compared to the size of the monkey legion. How is it possible that a vanishingly small number of cells in Shakespeare's brain managed to produce his entire literary output within a vanishingly small time period (compared to the age of the universe)? The monkey legion is utterly disorganized while the activity of brain cells is precisely orchestrated, what are the

principles and mechanisms of such orchestration responsible for the staggering difference in the output? Taking a closer look at the construction of Shakespeare's texts might offer some clues about the organization of brain processes.

Shakespeare's complete works comprise 884,647 words arranged in 118, 406 lines. Applying statistical measures, one can find out, for example, that predictability of letters (entropy per letter) in Shakespeare's texts depends strongly on the letter's position in the word, declining from roughly 3.8 bits in the first letter to 2 bits in the second letter and reaching a plateau of 0.7 bits after the fifth letter. These statistical characteristics are not particularly informative since they do not change much when the words are randomly scrambled, nor there is much difference between Shakespeare's text and a collection of mixed English texts from newspapers [16]. More sophisticated methods of text analysis apply measures of information-based energy and (information-based) temperature to detect variations in the text organization (words with different occurrence frequencies are placed at different energy levels presumed to obey Boltzmann distribution, and the relative temperature of a selected piece of text is computed as the ratio of energy measures in that piece and in the entire corpus). When applied to the collection of Shakespeare's plays, the method revealed that, among the four genres (histories, comedies, tragedies and romances), tragedies have the highest relative temperature (histories have the lowest) and *The Tragedy of Macbeth* scores the highest among the tragedies [17]. How so?

Study in [17] interprets relative temperature as a characteristic of the author's ability to choose words and construct texts in a manner that is both succinct and gives the fullest possible expression to the underlying thoughts (manifesting most prominently in *Macbeth*). Presumably, exercising this ability in the production of literary works (e.g., writing plays) is aimed at maximizing understandability, that is, affording readers the best means for understanding the author's thoughts and intentions. Understandability can serve as a decisive criteria in assessing differences between Shakespeare's texts and monkeys' output: the overwhelming bulk of monkeys' production is gibberish while Shakespeare's works are understandable and profoundly meaningful.

Per Webster's definition, text understandability depends on the extent to which the selection and composition of words are conducive to a) expressing relations considered by the author and b) constructing relations in the reader's mind isomorphic to those entertained by the author. What is unique about *Macbeth* that could both make the play particularly understandable and also account for the results of statistical analysis? Consider three lines at the apex of the play (scene 23):

Seyton. *The queen, my lord, is dead.*

Macbeth. *She should have died hereafter,*

There would have been time for such a word...

The last two lines present the entirety of Macbeth's reference to the queen in his response to the tragic news; made on the eve of the decisive battle, they convey, in the most succinct and powerful manner, the feeling of despair and a foreboding of the forthcoming military defeat. By wishing to shift the sad news to the "hereafter", Macbeth assigns it a level of significance no lesser than that of the expected military rout and his own likely demise, thus conveying the feeling of a total catastrophe without making any verbose statements to that effect. The following observation concerns a feature of understanding capacity that is presumed to manifest prominently in the cited text, and will play a pivotal role in the theory of understanding outlined in the subsequent sections. Observe that, when constructing the plot, Shakespeare was free to invoke the queens' departure at any point, including allowing her to outlive her husband. The exact timing, neither a day earlier nor at

any time “hereafter”, must have been decided from the start precisely to motivate the striking expression of despair and the subsequent monolog which expanded the meaning of the play from a chronicle of particular (imaginary) events to a philosophical generalization concerning the inescapable drama of the human condition. The monolog starts with the two lines above and concludes with some of the most quoted passages in Shakespeare’s literary legacy.

*“Life’s but a walking shadow; a poor player
That struts and frets his hour upon the stage,
And then is heard no more: it is a tale
Told by an idiot, full of sound and fury,
Signifying nothing.”*

The *Tragedy of Macbeth* involves 31 personages, including witches and apparitions, acting in small groupings in 25 consecutive scenes, as shown in **Figure 1**. The sparse matrix in **Figure 1** reveals the overall organization of the play emanating from the organization of the author’s mental model that, presumably, formed at the conception of the play and controlled its unfolding.

To underscore, **Figure 1** connotes that, in the mental model, interactions between personages are neither serial nor parallel but simultaneous (or “co-instantaneously co-ordinated”, as termed by Jean Piaget in [18]). For example, in scene 3, witches prophesize to Macbeth which results in changing the state of his mind; in scene 7, Macbeth influenced by the prophecy kills Duncan, which was made possible by Duncan’s arrival in Macbeth’s castle in scene 6, etc. Macbeth’s monolog expresses Shakespeare’s pessimistic worldview that is echoed in his other plays, for example:

Prospero. *“...Yea, all which is inherit, shall dissolve,
And like this unsubstantial pageant faded,
Leave not a rack behind, we are such stuff
As dreams are made of, and our little life
Is rounded with a sleep.”* *The Tempest, Act 4, scene 1*

Arguably, the corpus of Shakespeare’s work, i.e. all 884,647 words in 118, 406 lines, is a congruent expression of a worldview rendering all human affairs, excepting those serving the basic survival needs, both superfluous (see the epigraph to this chapter) and devoid of significance. It appears that exceptionally tight action coordination in the plot of *Macbeth* combined with succinct expressions of the

	Scene 1		Scene 3		Scene 24	Scene 25
1. Macbeth			1		1	1
2. Lady Macbeth						
3. Macduff					1	
4. Duncan						
5. Malcolm						1
.....						
29. First Witch	1		1			
30. Second Witch	1		1			
31. Third Witch	1		1			

Figure 1. *Macbeth plot comprises tightly coordinated interactions among numerous personages and unfolds in consecutive scenes each involving a subset of personages.*

author’s worldview consistent with other such expressions throughout the corpus have surfaced in the text features detected by statistical measures [17].

To summarize what has been suggested up to this point: “apprehending general relations in a multitude of particulars” (per the definition in Webster’s Dictionary) takes the form of constructing simultaneous memory structures where entities and their behavior are tightly coordinated. “Relations” are different forms of behavior coordination, i.e., a particular manner in which changes in one entity entail changes in other entities. When entities admit multiple states and a variety of state transitions, relations determine particular mappings between state transition sequences (state trajectories), as shown in **Figure 2**.

Models can form hierarchies where relations in the upper-level models (general relations) admit different instantiations in the lower levels (e.g., a worldview instantiated in different plays). Mental modeling enables predictions, explanations and planning under unfamiliar conditions, by ‘running’ models to generate predictions and then using predictions to inform the responses. These functions are made possible by coordinations preventing combinatorial explosion that would have made them intractable. One more literary example (adopted from [19]) will help illustrating these important notions.

Two elderly gentlemen, A and B, are waiting together for a train when a presentable looking young man (C) approaches A, politely asking for the time. After a short glance at C, A curtly tells C to leave them alone. When confronted by B about the rude response, A explains: “I thought that if I answered this young man, he might stay with us and keep the conversation going – next, he might board the train with us - next, he might get off the train with us – next, it might happen that my daughter D will come to meet me at the station – next, my daughter and the young man might like each other – next, they might start dating and will eventually marry – next, my daughter might end up unhappy because she married a man who can’t even buy himself a watch.”

Note that the model was a) composed on the spot to account for a peculiar set of circumstances (as opposed to being retrieved by matching or forged by filling slots in some pre-fabricated template), b) included a chain of tightly coordinated components connecting current conditions to their likely remote consequences (the prediction) and c) enabled using predictions to form a response deviating sharply from the habitual pattern (i.e., rude response to a polite question). More precisely, the model formed by A is a composition of globally coordinated and tightly constrained activities (e.g., C could choose any spot on the platform but was pinned down to the vicinity of A and B, he could board any train and get off anywhere but was constrained to follow A and B, daughter D could be doing anything anywhere but was constrained to appear at the railway station at the time of train’s arrival, etc.). The model instantiates a general relation (between income and matrimonial success) held by A, enabling him to predict events in the distant future (D will be unhappy) based on the current cue (C has no watch), and then to use this prediction to inform the immediate response and to explain the prediction and the response to B.

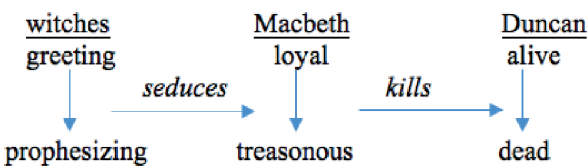


Figure 2. Relations establish coordinations between state trajectories. Models are simultaneous structures coordinating deployment of relations and self-initiated state changes (e.g., deploying relation “Macbeth kills Duncan” is preceded by Duncan’s decision to put himself in the harm’s way, by visiting Macbeth’s castle).

Importantly, models admit deliberately inserted counterfactual variations (e.g. A could have second thoughts and imagine C owning an expensive watch and asking for time because it had accidentally stopped) and generate the corresponding predictions (e.g., a satisfactory marriage) without revisiting the path (i.e., skipping over the sequence “C will stay with us, board the train with us, etc.)). Similarly, the model allows assessing global impact of local changes in one of the components without giving consideration to other components (e.g., one does need to trace the chain of coordinations in order to realize that failure in one element (e.g., D does not come to the station) will fail the entire chain and cancel the prediction). Crucially, models ‘resist’ relaxation of constraints, requiring forceful (deliberate) insertion of variations (i.e., under the model, the thoughts of D failing to appear, or C owning an expensive watch, etc. do not come to mind, as opposed to being rejected upon examination).

To summarize, eliminating degrees of freedom in mental models entails removing from consideration an otherwise exploding multitude of alternatives, thus making predictions both attainable and usable (i.e., delivered within the time window demanded by the situation). Models admit local variations consistent across the model (e.g. Macbeth’s decision to kill Duncan in scene 7 is consistent with what happened to him in scene 2, etc.) and suppress spurious variations. As a result, understanding yields the experience of having succeeded in grasping “general relations in multitudes of particulars”, thus turning an intractable mess into a well ordered structure.

The next section turns from literary scenarios to realistic ones, seeking to illustrate the extremes (amazing successes and baffling failures) in the operation of understanding.

2.2 From children’s games to revolutionary discoveries

2.2.1 Baffling failures

Children at an early age often fail to connect and coordinate events taking place right in front of them, as follows. The child is shown a toy which is subsequently placed under a cover allowing her to retrieve the toy. After a few successful repetitions, the toy is transferred, in full view of the child, to another spot where it is placed under another cover. After some hesitation, the child looks for the toy, not in the spot to where it was just moved but in the previous one [20].

Claudius Galen, an outstanding philosopher and physician in the Roman Empire, formulated a theory of blood production and processing in the body (circa 150 AD). The theory asserted that blood is produced in the liver from ingested food, rises to the lungs through the right side of the heart, crosses through pores to the left side where it is mixed with inhaled air and, finally, gets distributed throughout the body and consumed by the tissue (the surplus is expelled with sweat and urine). In this schema, heart remains a reservoir where blood is collected and treated (mixed with air) on its way from the source (liver) to the sink (tissues). In the XIth century, Galen’s works were translated into Latin and became a dogma that dominated medical profession for over 500 years. Ironically, bloodletting was one of the most frequent treatment modalities in the medieval medicine, but neither the viewing of blood streams spurting from incisions nor the evidence of heart’s incessant beating in one’s own chest could cause questioning of the dogma. In 1628, English physician William Harvey published a book presenting a simple and cogently argued model of blood circulation. Moreover, he pointed out absurdities inherent in the dogma (e.g., the liver would have to produce several times the body weight in blood every day if the blood was being absorbed). Despite their undeniable strength (a simple model

accounting fully of the available data and revealing critical shortcomings in the earlier account), Harvey's ideas were met with ridicule [21, 22]. The medical profession was unable to overcome the inertia and re-structure the entrenched model, thus failing to apprehend coordination between a few vital variables. Galen was an expert on pulse diagnosis and published a treatise on the subject, which makes his conceptual blind spots particularly baffling. Perpetuation of Galen's model would have arrested progress in medicine, causing incalculable losses (think of Galenic cardiology).

2.2.2 Spectacular successes

In the 1820–1835 time period, Michael Faraday formulated key ideas of the field theory postulating relations between electric and magnetic phenomena which, in the preceding decades, were commonly viewed as being totally unrelated. Expressed in a mathematical formalism by James Clerk Maxwell, the Faraday - Maxwell model of electromagnetism depicted propagation of electric and magnetic fields as tightly coordinated processes. Faraday's conceptualization of fields envisioned material entities of a kind that are not perceptually accessible but permeate space and carry force. In a brilliant feat of expansive insight, Maxwell realized the existence of relations between electromagnetic waves, light and perception of color. These findings have been propelling advances in physics and technology, until the present day and into the foreseeable future.

Modern physics (quantum mechanics, astrophysics) deals with entities that are not directly observable. Literature reports that key ideas concerning quantum processes were formulated by Werner Heisenberg (circa 1925) following an insight he allegedly received when taking a walk in the park at night and observing a passer by appearing in illuminated areas under lamp posts and disappearing in the shadows when leaving those areas [23]. The position and movement of the person between the posts remained undetermined, suggesting the idea of indeterminate states of electrons in the atom when transiting between energy levels (somewhat similar to indeterminate states of characters in a play when transiting between scenes, as in **Figure 1**). Quantum mechanics proved to be the most successful physical theory ever formulated, predicting the outcomes of particle interactions with unparalleled accuracy.

As reported in [24], an explosion on a DC-10 passenger airliner incapacitated one of three engines and demolished the hydraulic system, causing loss of control mechanisms for the remaining two engines except for their thrust levers. Hydraulic systems are built with triple redundancy, bringing the odds of losing control due to hydraulic system failure to less than one in a billion. Accordingly, no protocol has been ever created for handling such occasions and no training was ever offered. When the aircraft started pitching violently up and down (a phugoid pattern), the pilot had a short time window to figure out how to suppress phugoids and land the aircraft. According to pilot's recollections, a simplified model was formed in his mind that accounted for the location of the remaining two engines and suggested a maneuvering strategy using differential thrust. The strategy was not only unfamiliar but grossly counterintuitive, requiring decelerating when the aircraft was climbing and accelerating when it was heading down. When flight conditions were reproduced in a simulator, numerous pilots failed to figuring out a course of action and kept crashing (could not make the runway after dozens of attempts) [24].

Samuel Reshevsky, a chess prodigy born in 1911 in Poland, learned the game at the age of four and at the age of eight was defeating champions of his country in tournaments, as well as beating scores of opponents, including master-level players, in public demonstrations of simultaneous play. Although cognitive difficulties faced

in chess have been always appreciated, there were no satisfactory methods for quantifying them until the era of chess computers. Chess algorithms required hardware with operating speed at or above 10^8 position evaluations per second in order to compete with expert players capable of carrying out at most one or two position evaluations per second. Understanding the game compensates for the $1:10^8$ disadvantage in speed: expert players perceive configurations of pieces as compositions of “complexes”, deriving game plans from apprehending coordinations between the “complexes” [25]. Findings in [25] suggest that expert game models take the form of simultaneous structures, not unlike the matrix in **Figure 1**. A novice’s perception is limited to a few adjacent cells in the matrix (2–3 moves look-ahead involving 2–3 pieces) while expert models can include a hierarchy of matrices encompassing the entire configuration and extending to 10–15 moves look-ahead. Position analysis involves envisioning variations for some of the moves, constrained by the entire web of coordinations across the matrix. As a result, experts are not distracted into considering spurious (weak) moves, no more than novices waste effort in considering illegal moves [26].

To summarize, the previous section associated understanding with the development of mental models representing entities, their behavior and different forms of behavior coordination in the form of simultaneous memory structures. It was suggested that simultaneous coordination suppresses combinatorial explosion, confining the process to an infinitesimally small volume in the vast combinatorial space (considering possible move combinations in chess, similar to considering possible letter combinations in playwriting, quickly brings one to the realm of counting protons in multiple universes). Prediction, explanation and planning are enabled by mental modeling. This section reviewed extreme cases when modeling processes failed to establish coordination between a few directly observable and persistent entities and succeeded in quickly coordinating multiple, transient and/or unobservable ones. Summarily, suggestions and observations in Section 2 define the main challenges facing a theory of understanding:

- a. what neuronal mechanisms can account for the successes and shortcomings of the understanding capacity,
- b. how such mechanisms could emerge and
- c. how could they develop in the human species within the time period of negligible duration (on the evolutionary time scale).

The next part focuses on the emergence of understanding.

3. A brief history of understanding

Notions addressed in this part were developed elsewhere [10–13, 27–30] and will be summarized briefly here. A preview will help putting the notions together: Environment is in flux, survival depends on an organism’s ability to adapt to the changing environment. Adaptation makes the world livable while understanding makes it intelligible, that is, amenable to prediction and explanation (i.e., connecting likely future events to their plausible causes in the past and present). Mechanisms of understanding complete the transformation of sensory streams into world models that generate such predictions and explanations. The transformation starts with mechanisms of sensation and perception that are available, in different forms, in other species, and culminates in the mechanism of understanding unique

to humans. Learning response-reward (response-punishment) patterns increases reward chances and decreases punishment risks when conditions recur. A repertoire of such learned patterns constitutes a model of the environment instantiated by pattern matching. Understanding is an advanced adaptive mechanism serving to overcome the inertia of prior learning and optimize responses when conditions are novel or violate the previously acquired conditions-response associations in a consequential manner (e.g., learned responses cease to be rewarding) [10–13]. This characterization is consistent with definitions of intelligence in the literature (“fluid intelligence” [18, 31–33]) establishing understanding capacity as the central, defining feature of human intellect.

3.1 Evolutionary precursors

Complex life forms have been developing on Earth at an accelerating pace: From the emergence of unicellular organisms some 3.7 billion years ago, to (the emergence of) multicellular animals 900 million years ago, to vertebrate 530 million years ago, to primates about 70 million years ago, to the detachment of the human branch from the chimpanzee/bonobos primate branch 6 million years ago to, finally, the emergence of anatomically modern Sapiens [34] at the time period of 200,000–100,000 years ago (the emergence of language is attributed to the time period of roughly 150,000–60,000 years ago [35–37]).

Recent findings indicated genealogical continuity in Sapience in the last 28,000 years, i.e. from Upper Paleolithic to modern times [38]. During the same period, the size of the braincase has been decreasing, having lost more than 10% of its peak value [39]), after a preceding period of about 6 million years during which the size almost tripled [40]. Recent analysis comparing the results of electrophysiological, anatomical and fMRI studies in humans and non-human primates associated development of intelligence primarily with reorganization of brain mechanisms [41]. These findings seem to indicate that reorganizations entailed higher efficiency so that progressively more complex tasks could be carried out

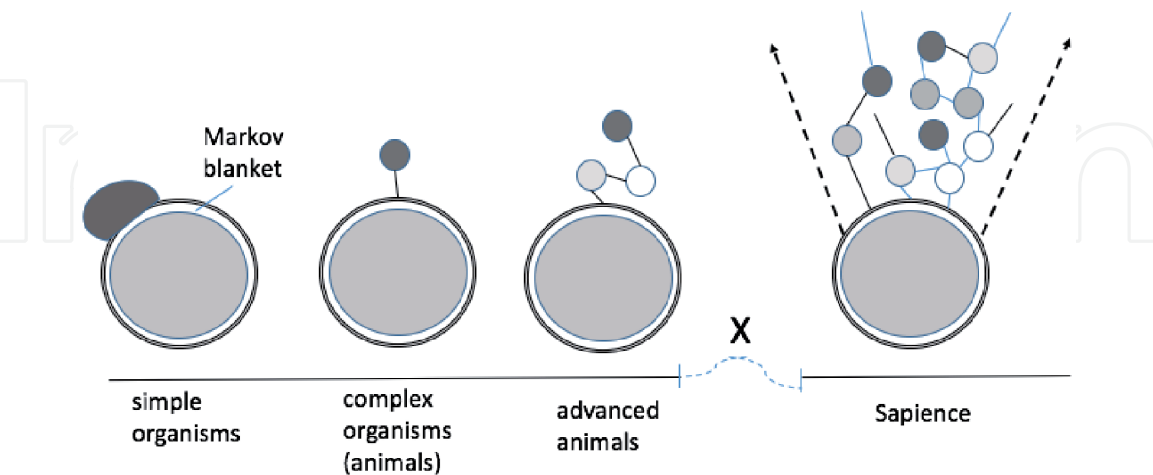


Figure 3.
Gap X denotes discontinuity in the development of cognitive capacities. Simple organisms interact with substances located on their ‘blankets’, more complex organisms can move towards and reach for target objects (denoted by black circles) in close proximity to their blankets (e.g. salamanders shoot their tongues to catch insects), and advanced animals (apes, some avians) can use a few supplementary objects (denoted by shaded and white circles) to act on the target (e.g. chimpanzees can connect sticks and pile up boxes in order to reach a hanging fruit). Humans are discontinuous with the other species in that they can form coordinated structures (designs) comprising indefinitely large sets of supplementary objects giving access to indefinitely distant targets, with the possibility of postponing acting on such targets until some indefinitely remote future moments (anticipatory planning).

without increasing the size of the neuronal pool. Section 4 will suggest the type of reorganization that could produce such revolutionary improvements.

Comparing modes of interaction between the organism and environment across the spectrum of life forms reveals discontinuities between Sapiens and other species, as shown in **Figure 3**. The term ‘Markov blanket’ [13, 14] denotes an enclosing boundary (e.g., membrane) separating organism from the environment (the notion will be defined more precisely in the next section).

Differences between Sapience and other species are qualitative: they lie not in the increased quantity of supplementary objects but in the drive to keep extending the reach of action (action envelope) and to form progressively more complex designs comprising growing numbers of objects of increasing variety. Stated differently, animal envelopes are limited to the immediate proximity of their Markov blankets while human envelopes undergo indefinite expansion. Amplifying Shakespeare’s insight (expressed succinctly in the epigraph), it can be suggested that animals seek biological equilibrium with their environment (i.e., maintaining inflows of energy and nutrients at life-sustaining levels) while humans seek cognitive equilibrium entailing demands not reducible to those associated with sustaining life. Hence, gap X. What is the nature of that gap?

3.1.1 Learning and pattern recognition

Consider challenges facing organisms in a changing environment. Assume first that the varying flow of conditions (stimuli) includes some recurring patterns. Since finding successful responses consumes time and effort, recognizing such patterns and re-using the responses saves both. The strategy works best when patterns comprise a few contiguous stimuli that trigger a small repertoire of fixed responses. However, even this simple strategy working under favorable circumstances can become self-defeating when the circumstance change, as illustrated in the following example.

Salamanders shoot their tongues at objects (insects) whose size, speed and distance from the animal fall within some fixed ranges, which requires anticipatory response control (early activation of the projector muscle relative to the tongue launch) to improve the chances of successful intercepts. The shooting mechanism was fine-tuned by evolution (developing spring-loaded type of tongue ejection yielding high energy output), making the animal a successful predator [42]. Consider a hypothetical scenario when the advantages are turned into detriments. The shooting mechanism is thermally sensitive: the speed of tongue retraction increases with temperature [42] which can be used, potentially, to increase the amount of prey intake per unit time. Assume that the animal can learn the ‘higher temperature – higher intake’ association, compelling it to seek high temperature spots. Such learning will keep paying off for as long as the prey cooperates: if the insects start moving faster in the vicinity of hot spots (or avoid them, etc.), the intercept success rate will decline. However, the animal will be bound to continue the heat-seeking behavior until the association decays, which might cause it to die from hunger and/or exhaustion (missing targets decreases food intake but not the costs). The point is that the ability to suppress learned behavior can yield quantum leap improvements in adaptive robustness, by reducing the probability of ‘blind persistence’ types of error inherent in recognition-centered strategy, and/or reducing the severity of the consequences. In general, the strategy works if short contiguous patterns (compact patterns) recur with frequency sufficient for satisfying the organism’s survival needs. Assume that the requirement is not met, forcing the animal to seek strategies applicable in more complex stimuli configurations.

3.1.2 Gap X

Removing (or relaxing) the contiguity requirement changes an animal’s view of the environment: from a noisy stream of compact patterns to a stream of uncertain structure where patterns can no longer be readily discerned. Stated differently, in streams of non-contiguous patterns (dispersed patterns) stimuli groupings in one pattern can be interspersed irregularly with groupings belonging to other patterns, thus allowing extending patterns over indefinitely long stimuli sequences and time periods. Dispersed patterns place organisms at the horns of a dilemma, as shown below.

Pattern composition in **Figure 4(2)** is inherently uncertain, gradual reduction of the uncertainty proceeds reversibly through the stages of a) defining entities (as compositions of states), b) defining behaviors (as patterns of state transition) c) defining relations (as forms of behavior coordination), resulting in the construction of simultaneous structures representing interactions between entities in successions of episodes, as shown in **Figure 5**.

Strategies in **Figure 4(1)** and **(2)** reside at the opposite sides of gap X: cognitive operations underlying the former are exogenously driven, i.e., triggered by the environment and carried out under feedback control, while operations underlying the latter are endogenously-driven, i.e. decoupled from the sensory inflows. Rudimentary forms of such decoupling manifest in animal behavior, e.g., dogs following a prey that disappears behind an obstacle might not chase it around the corner but run to intercept at the opposite corner. On the human side of the gap, reversible operations become available gradually as the person matures, causing characteristic errors (e.g., young children fail in the “toy has moved” task requiring that association (toy, cover1, spot1) is followed by dissociation (toy, cover1, spot1) → (cover 1, spot 1) → (toy, cover2, spot2), see section I.2.a. A different form of dissociation deficit manifests in older children when they fail to dissociate container from the contents: a child watching liquid being poured from one container to another can believe that the amount changes with the size of the container [43].

Note that entity construction principles in **Figure 4(2)** and **5** express an implicit assumption that entity’s identity can be preserved in different manifestations in non-contiguous episodes, that is, the same entity can have different (non-overlapping) manifestations and, vice versa, different entities can have identical manifestations (e.g., in Greek mythology, enterprising Zeus was appearing to mortal women in the form of a swan, a bull, or even a shower. On one occasion, Zeus presented himself to a lady in a form that was identical to her husband (Amphitryon) in every

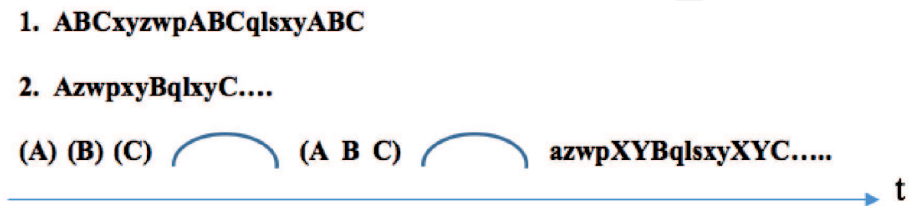


Figure 4. Transition from compact to dispersed patterns inside gap X. 1) contiguous stimuli grouping ABC recurs at irregular noisy intervals, response strategy consists in finding activities rewarded by ABC and emitting them whenever the pattern is recognized. 2) removing the contiguity requirement changes the strategy from pattern recognition to pattern construction. Here is the dilemma: stimuli A, B, C can be manifestations of either different entities requiring different responses or different states of the same entity requiring the same response (possibly, with modifications). Whatever the resolution, it might change at some later point in time, e.g. XYB and XYC can be determined to be the states of some entity Z, causing a to recede into the background noise, etc.

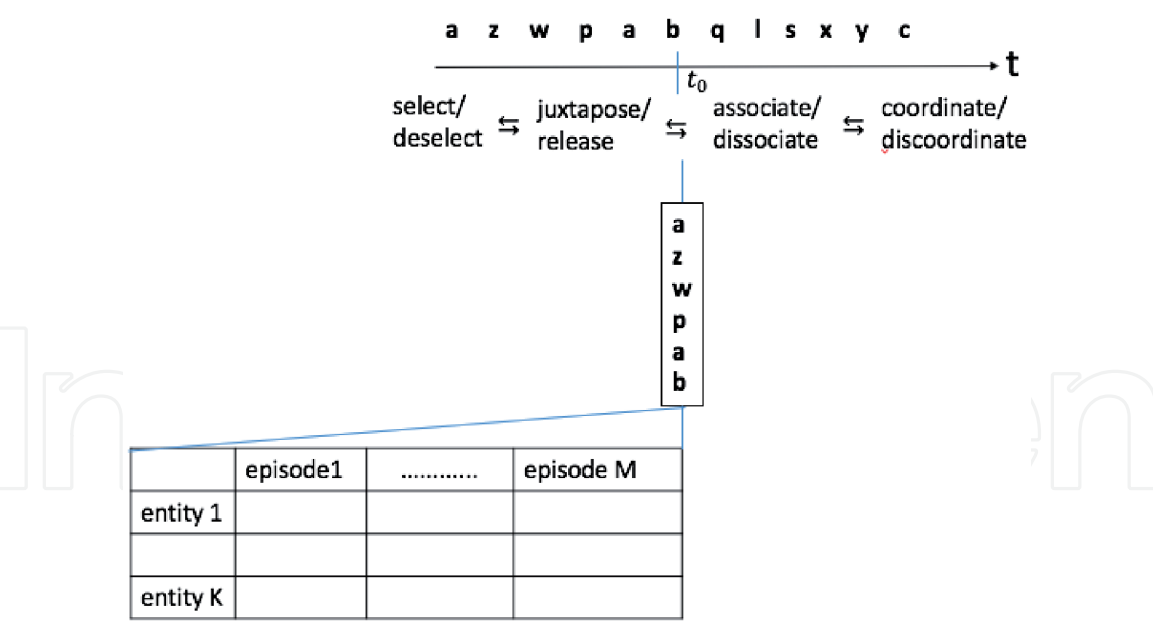


Figure 5. An irreversible stimuli stream is transformed into a simultaneous record, cycles of reversible operations on the record (select/deselect, etc.) produce simultaneous structures comprising various entities interacting in series of episodes (see **Figure 1**).

detail but was not her husband – Amphitryon was quite sure of that). Implicit explorations of logic in Greek mythology were made explicit by Aristotle in the Laws of Thought, including the Law of Identity.

3.2 Crossing gap X

According to an appealing hypothesis [44], the earliest steps in the expansion of the human envelope were associated with predation by throwing projectiles (stones). Accurate aiming requires precise coordination of several variables including launch angle, velocity, weight and size of the stone, distance to the prey and its size, and release time, with the width of the release time window limited to a few milliseconds (e.g., 11 milliseconds for a rabbit-size stationary target located 4 meters away, these results will be re-visited in the next section). Analysis based on experimental findings (narrowing the time window involves synchronization in neuronal clusters of growing size) demonstrated that increasing distance to targets while maintaining the hit rate requires explosive growth in the number of neurons responsible for precise timing (64-fold and 729-fold increase in the number of neurons to double and triple the distance, correspondingly).

Anatomical limitations imposed on the volume of cranial cavity appeared to exclude the possibility that a growing variety of high-precision activities (e.g. splitting stones for different tasks) could be obtained by developing narrowly specialized neuronal modules. Anatomical limitations enforce other trade-offs having impact on cognitive performance, e.g., increasing the speed of pulse conduction would require increasing the thickness of myelin wrappings, which would decrease the number of neurons the cranial geometry can accommodate [40, 44] .

In addition to constraints in brain size and conduction speed, another physical factor having decisive impact on brain processes is limited supply of energy for powering them. Since physical constraints on brain processes are non-negotiable, the only avenue for obtaining quantum advancements in cognitive performance depicted in **Figure 2** appears to be dynamic optimization in their deployment, which boils down to global coordination via the mechanisms of mental modeling.

These notions will be addressed in the theory of understanding in the next part, following another example of mental modeling in the closing of this part.

For the sake of argument, assume that advancing the predation-by-throwing-projectiles strategy involved invention of catapults, in the simplest form of a board (B1) balanced on a base, or fulcrum (B2). Note that neither component, if considered individually, betrays any hint as to its potential usefulness for projectile throwing. Moreover, when considered jointly, these components afford numerous arrangements that are all useless (e.g., the base on top of the board, etc.), with only one particular form of base-board position coordination yielding the benefit. Operations involved in constructing and operating catapults are suggested in **Figure 6**.

Note that the product of modeling is a new entity (a weapon) that has properties unavailable in the components and expands the activity envelope (larger distances, heavier projectiles). Running the model yields understanding, i.e., informs operation and aiming procedures. For example, envisioning one side of the board going up brings to mind the image of the other side going down, envisioning increasing the distance to the target brings to mind the image of increasing the length of the shoulder (shifting the projectile away from the base), etc. That same process underlies prediction (e.g. hit probability) and explanation (why hitting that target over there is unlikely?).

It is interesting to note that children up to a certain age, when learning to operate toy catapults, are often incapable of forming proper models and keep shifting projectiles in the direction of the target as it moves away (shortening the shoulder), even after having watched the proper operation multiple times [43]. The instinctual tendency to grasp receding objects by extending arms and moving after the objects resists learning. Young children cannot understand catapults.

Recent theories concerning the origins of language placed the capacity to perform reversible juxtaposition (operation Merge) at the foundation on which all other language mechanisms have been built (B2 B1) → C (operation Merge combines syntactic objects in an arbitrary order [45]).

To summarize, this part defined understanding as an advanced adaptive mechanism that makes possible constructing responses to indefinitely large patterns comprised of non-contiguous stimuli groupings (dispersed patterns). Construction proceeds through identifying entities, their properties and behavior and the forms of inter-entity behavior coordination, culminating in the production of simultaneous, tightly coordinated structures comprising multiple entities (mental models). Models are amenable to manipulations, giving rise to the dual capacity for predicting likely events (changes in the entities) and identifying their causes in the past or present (explanations). In general, any organism can be viewed as a cast molded by the environmental niche it occupies, e.g., salamander is a ‘cast mold’ of environment where particular (edible) insects having size and speed within some fixed ranges are flowing into a volume in space reachable by the animal in unit time in quantities sufficient for the animal’s survival. The total model includes biophysical component

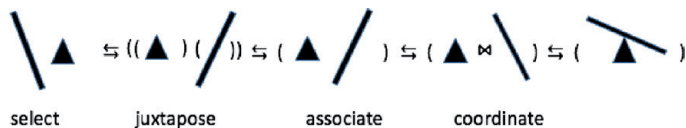


Figure 6. Modeling starts with selecting entities (objects) and juxtaposing them as separate (independent) entities, followed by associating them in a composite structure allowing inter-dependence, followed by coordinating the entities to form a model (note that juxtaposition brings components together in an arbitrary order while association imposes order, setting the stage for establishing a higher degree of order in the model). Symbol $\times$ denotes coordination.

(body and the sensory-motor periphery, e.g. the tongue-ejecting mechanism) and regulatory component orchestrating activities within the body and at the periphery (i.e., animal's behavior in the environment). Both components undergo evolutionary development in the species while behavior regulation is amenable to adaptive changes in individuals during their lifetime (learning). In animals, learning is restricted to condition-driven variations within narrow envelopes of genetically-fixed condition-response patterns and propensities. Condition-driven learning extrapolates from past precedents while mental modeling enables prediction and response construction under conditions having no such precedents. More precisely, models integrate past history within cross-coordinated structures so predictions produced by operations on the structure can be made consistent with (plausible under the entire past history) without repeating any of its elements. Moreover, models allow reproductive construction without replication, e.g., coordinations in the basic catapult were reproduced in numerous designs.

As observed by Jean Piaget [46].

"...mental coordinations succeed in combining all the multifarious data and successive data into an overall, simultaneous picture, which vastly multiplies their powers of spatio-temporal extension, and of deducing possible developments"
([46] p. 218).

Summarily, it has been suggested that a) the protohuman-to-human transition was associated with the emergent capacity to construct responses to dispersed stimuli patterns and b) the capacity is rooted in the mechanisms of mental modeling that represents such patterns as coordinated structures that suppress combinatorial explosion inherent in the construction process and reduce the number of response compositions to a few plausible alternatives.

4. Theory of understanding: neuronal mechanisms of mental modeling

The theory in a nutshell: Nervous system optimizes deployment of sensory-motor resources vis-à-vis varying external conditions, a part of the system that coordinates variations in the deployment of sensory motor-resources with variations in the conditions flow constitutes the first regulatory loop. Mechanisms of understanding operate on top of the first loop and optimize the organization of neuronal resources engaged in that loop, thus forming the second regulatory loop. Optimization in the second loop involves arranging neurons into coordinated structures manifested in coordinated mental models, as shown in **Figure 4**. Operations in the first loop are controlled by sensory-motor feedback while operations in the second one are decoupled from it. Feedback control makes resource deployment adaptive, self-controlled optimization in the second loop makes it self-adaptive [11]. First and second loops are stages of self-organization in the neuronal substrate. The first loop allows adaptation to compact sensory patterns extending over short time periods while the second one expands adaptation to dispersed patterns extending over indefinitely large time periods (prediction). Limitations on the size of the neuronal pool and the amount of usable energy supplied per unit time drive the need to increase adaptation span while reducing energy costs, which boils down to a dual optimization criteria: minimize energy losses and the amount of energy consuming activities while maximizing prediction accuracy. Both criteria are subsumed under the notion of active inference [13, 14, 47].

This part will discuss the role of understanding capacity within the active inference framework, followed by detailed suggestions regarding neuronal mechanisms

that underlie the capacity and are responsible for the range of its operation, including the extremes. The part concludes by referencing experimental findings and ideas in the literature that might help in assessing biological plausibility of the present proposal.

4.1 Active inference: from Aristotle to Friston

The opening line in Aristotle's *Metaphysics* states that "humans desire to understand" [48]. Lack of understanding engenders puzzlement, and failure to identify causes leads to undesirable self-evaluation

".. men of experience know that thing is so, but do not know why, while the others know the 'why' and the cause.... and man who is puzzled and wonders thinks himself ignorant" (Aristotle, Metaphysics).

In a penetrating insight, Aristotle captures relations between experiences, surprise (puzzlement) and self-directed activities motivated by the desire to reach beyond the appearances (identify causes). Arguably, principles of active inference and variational free energy minimization advanced in [13] are congruent with those early insights. The principles assert that life in all its forms, from unicellular organisms to humans, is predicated on the organisms ability to use sensing to predict conditions in its environment and to conduct activities reducing the difference between the predicted and the actual experiences. Predictions require models of the environment, the variational free energy value determines, roughly, the (information-theoretic) distance between the current and the desired states that takes into account the difference between the predicted conditions and those that were actually sensed and the surprise experienced under the model (the smaller the probability assigned by the model to the condition, the higher the surprise).

Emphasis on activities directed at minimizing variational free energy underlies the notion of 'active inference,' which is best appreciated if contrasted to the idea of 'passive' inference expressed in Plato's allegory of the cave, as follows. Prisoners are chained to the floor inside a cave where they can see nothing of the outside world except shadows on the wall they are facing. The message is that people are caged inside their minds, senses are the only window into the world, and that window can be distorting. The allegory defines passive inference: prisoners can make guesses about the outside world but have no means to validate them or to use in any fashion.

Active inference differs from passive inference in that it incorporates iterative actions on both the outside world and the model of that world that can lead to progressively improving guesses. Understanding involves a form of model manipulation that is best defined within the active inference theory through the notion of a Markov Blanket - the third conceptual pillar in the theory integrating ideas about emergence of life, evolution, and brain operation into a seamless whole.

'Markov Blanket of node x ' is a graph-theoretic term denoting a set of nodes in a directed graph that are connected to x by links incident to and from x . More loosely, the term can be used to denote a group of nodes in subnetwork X_1 separating it from the rest of the network X . If links denote some form of interaction, Markov Blanket of X_1 can be viewed as an interface through which internal nodes in X_1 interact with their surrounds in X . On that view, Markov Blanket accords X_1 a degree of (conditional) independence from X - a critical concept in the overall theory, as follows.

The theory of life attributes emergence of life to spontaneous phase transitions in molecular networks ('primordial soup'), resulting in the formation of subnetworks that remain connected to their surrounds but acquire a degree statistical

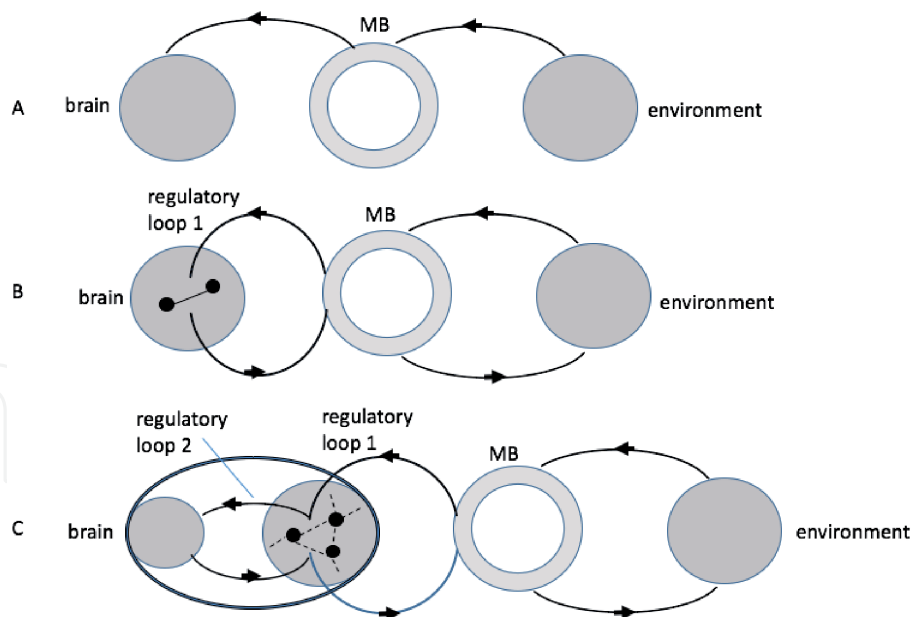


Figure 7.

Passive inference, active inference and comprehensive active inference. (A) The allegory of the cave (passive observation without action: sensory input is neither solicited nor acted upon). (B) Observation-action iterations guided by feedback produce (deposit) a model that adjusts subsequent iterations and gets adjusted by them. (C) Second regulatory loop manipulates structures formed by the first loop to construct models, the process is decoupled from the motor-sensory feedback.

independence (autonomy) from it. In that context, Markov Blanket denotes interface (a ‘membrane’) between such quasi-autonomous formations and their environment [14]. As more complex forms of life develop, the Markov Blanket expands to incorporate the entire sensory-motor periphery, as suggested in **Figure 3**. Finally, **Figure 7** separates Markov Blanket from the nervous system to illustrate the notions of active inference and comprehensive active inference (incorporating the understanding capacity).

The process in **Figure 7A** is an idealization; **Figure 7B** depicts associative learning (e.g. the hypothetical salamander associates elevated temperature with successful hunting, entailing search for hot spots); **Figure 7C** depicts active construction of mental models that underlies understanding. Learning yields “knowledge that a thing is so”, understanding defines causes.

In summary, different facets of the ideas depicted in **Figure 7** have been addressed in numerous sources in psychology, physiology, neuroscience and philosophy of the mind. The active inference framework offers a synthesis of some of the key insights in these disciplines, integrating them in a coordinated conceptual structure expressed in a unifying mathematical formalism. The central notion is that of activity: an organism is actively seeking sensory inputs, constructs models and acts on the environment. These contentions will be re-visited in the discussion.

4.2 Neuronal mechanisms

The proposal in this section stems from five assumptions about the nature of neuronal processes that underlie intelligence and its special form, understanding. The proposal will be presented in three sections: first, the assumptions are formulated, along with some clarifications; next, the key points in the theory are formulated and applied to answer questions posed at the end of Section 2; finally, these key points are re-visited and related to experimental findings and other ideas in the literature.

4.2.1 Assumptions

4.2.1.1 Cognition involves active deployment of neuronal resources

Brain is a synergistic system that selects, mobilizes and deploys (fires) neurons. Mobilization involves activities that precede firing and are centered on tuning, as shown **Figure 8**.

Consider the following three experiments: raising your right hand and touching your nose with the index finger, doing the same with your eyes closed, and imagining the same without doing anything. The first run involves coordination in the external space, the second involves coordination in the mental space (you know where your nose and your finger are, without reference to external coordinates), the third demonstrates coordination in the neuronal space that underlies the other two (I shall return to these exciting experiments at the end of the section).

4.2.1.2 Progressively improving deployment requires relative stability of neuronal groups

Deployment strategy progresses from deploying individual neurons to deploying neuronal groups, to deploying groups of groups, etc., which requires a degree of stability in all the elements of the growing organization. This intuition entailed the notion of “neuronal packets” that is pivotal in the theory.

A neuronal packet is Hebb’s assembly (i.e., comprises neurons connected by associative links) that is synergistic and is separated by a boundary energy barrier from the surrounding associative network.

It was hypothesized that packets form as a result of phase transition in associative networks, not unlike raindrops form in vapor. Accordingly, energy barrier is determined by surface tension, that is, the amount of free energy per unit surface (presumably, surface comprises cell membranes in the boundary neurons. Accordingly, surface energy is determined by the distribution of membrane potential across the surface). Neurons at the packet boundary constitute packet’s Markov Blanket, surface tension in the boundary holds neurons together. Mapping these notions on the process in **Figure 4** will help appreciating its crucial consequences: first, combining neurons responding to A, B, C, D, E... in a quasi-stable bounded packet amounts to asserting existence (perceiving) some bounded entity (object) α comprising features $\alpha = \{A, B, C, D, E...\}$ and, second, synergistic packets allow ‘tuning’ to their

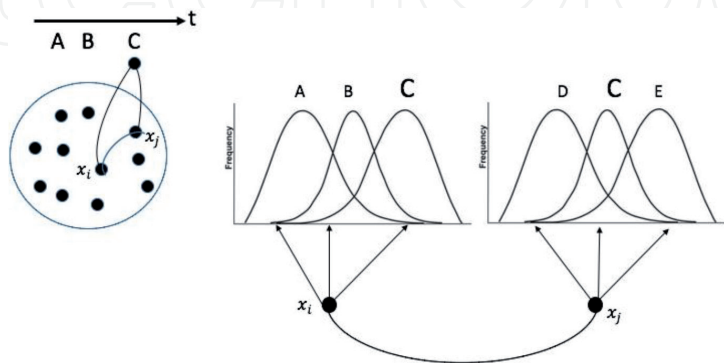


Figure 8. Neurons x_i and x_j are selected in the neuronal pool and tuned to stimulus C in the stimuli stream. Neuron x_i responds to A, B, C stimuli, tuning amplifies its response to C. Sensing and motor actions are both products of active deployment (e.g. one sees color C because some neurons were selected, mobilized and tuned to C). Imagining color C involves the same process. Imagining A, or B, or C involves shifts in tuning, which can be expressed as rotating neuron’s response vector. Co-firing of x_i and x_j establishes an associative link between them.

individual constituents (rotating packet vector) which is experienced as envisioning different states, or facets of object α (e.g., rotating the image). Energy barriers ‘anchor’ determinations in Figure 4, e.g., once feature A has been attributed to object α , the barriers will resist (require energy investment in) separating A from α . As a result, barriers serve the dual function of binding neurons together in stable groups and binding those groups to ‘objects.’ **Figure 9** illustrates these notions.

4.2.1.3 Improving deployment requires coordination of neuronal groups

Models are composite ‘objects,’ i.e., synergistic groups of coordinated packets. For example, neuronal group ‘catapult’ comprises packets ‘board,’ ‘base,’ ‘projectile’ and ‘target’ and can be ‘tuned’ to different states of the composite object. A crucial point: feature space of ‘catapult’ has dimensionality higher than that of the constituents, rotating the ‘catapult’ vector (e.g., switching between states ‘unloaded’ $\rightarrow$ ‘loaded’ $\rightarrow$ ‘aimed,’ etc.) reflects coordinated movement of the constituent vectors (e.g., envisioning a receding target brings to mind the image of a projectile moving away from the base). **Figure 10** maps these notions on the organization depicted in **Figure 7c**.

4.2.1.4 Brain is a self-organizing virtual system

Genetically-defined propensities in the brain substrate (gray and white matter, etc.) allow a range of self-organization trajectories, the actual developmental

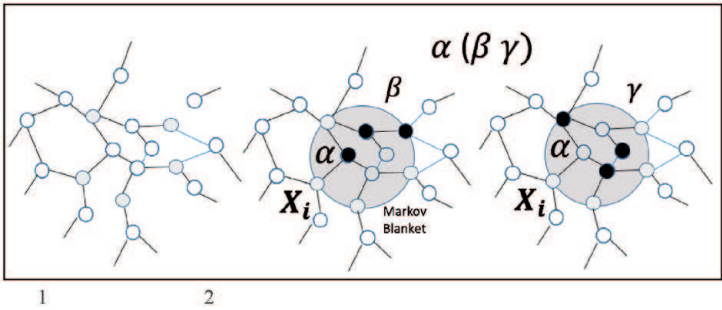


Figure 9. 1. Successive co-activation of different neurons produces a growing associative network. 2. Associative network undergoes phase transition resulting in the formation of packet X_i , giving rise to perceiving object α . Different activation–inhibition patterns in α underlie the experience of α manifesting states β and γ and behavior patterns) $\beta \rightarrow \gamma$ and $\gamma \rightarrow \beta$.

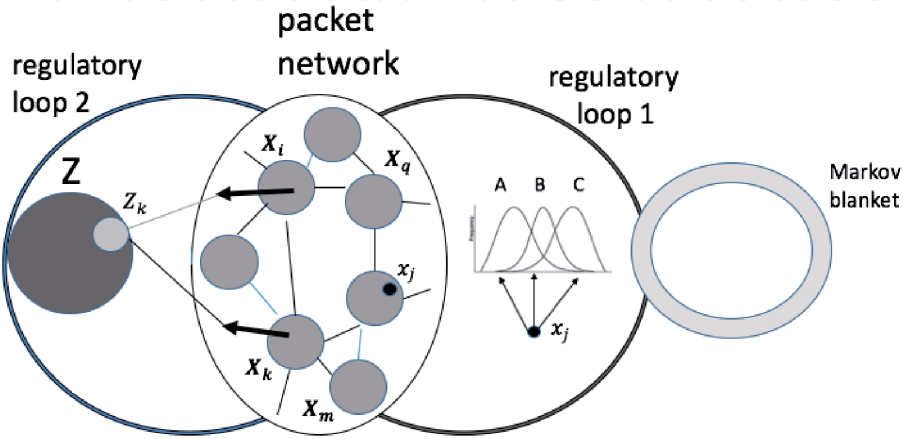


Figure 10. Phase transitions in the associative network transform it into a packet network. Selecting, mobilizing and deploying packets in the packet network populates the world with a multitude of distinct objects capable of different behavior patterns. Mental models establish coordination between behavior patterns.

trajectory results from an interplay between the propensities and conditions encountered throughout the lifetime.

4.2.1.5 Brain is an energy seeking system

Self-organization is predicated on energy inflows sufficient for producing coordinated neuronal structures. The process is sustainable because it stabilizes energy inflows via expanding the range of extremal activities (thus diversifying energy sources) while minimizing internal energy expenditures incurred in the expansion.

4.2.1.6 Self-organization proceeds through assimilation/accommodation cycles

Periods of deliberate (attentive, self-directed) construction and manipulation of mental models alternate with periods of spontaneous re-structuring: the overall neuronal organization adapts to the newly formed structures and, reciprocally, the new structure are adjusted and integrated into the organization.

4.2.1.7 Brain is a synergistic system

In neuronal structures, a few controls can manipulate a much larger numbers of degrees of freedom [49, 50]. **Figure 11** illustrates this important notion.

4.2.2 Putting it all together: neuronal substrate of understanding and brain functional architecture

Assumptions advanced in the preceding section entail the following suggestions.

1. Formation of neuronal packets transforms associative network into a packet network embedded into an energy landscape, with the packets residing in local minima. The height of packet energy barrier E_m (free energy) is a function of temperature T and parameter $\sigma(T)$ reflecting cumulative strength of associative links incident to the packet's Markov Blanket (MB) from inside the packet vs. the cumulative strength of those incident from the outside.

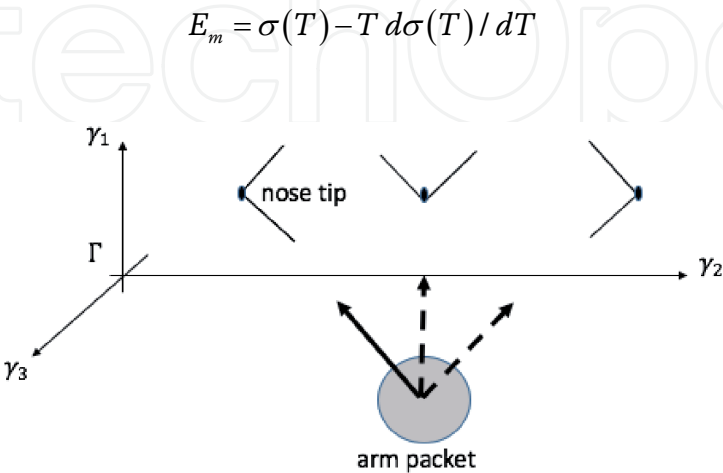


Figure 11. Imagine raising your arm and touching the tip of your nose in three consecutive positions: looking to the left, looking straight ahead, and looking to the right. Population vector in the packet determining arm movement rotates accordingly. Coordinates of the nose tip in mental space control tuning of numerous neurons in the arm packet. Seminal studies in [51–53] demonstrated that movement organization involves rotation of packet vectors in the direction of the target.

($\sigma(T)$) is analogous to membrane potential determined by the difference in ions concentration on both sides of the membrane, $\sigma(T)$ declines as temperature grows, E_m is an inverse of MB's permeability (resistance)). Packets connected by associative links might not be mutually accessible if separated by high energy barriers, as illustrated in **Figure 12**.

The height of energy barrier E_m determines relative stability of packet X_m that corresponds, roughly, to a level of subjective confidence in X_m , which can vary depending on the local temperature (the lower the temperature, the higher the barrier. Consistent with [54], temperature variations shape the landscape and facilitate jumps of free energy barriers. Under the notion that deployment of neuronal resources serves to extract free energy from the environment [11, 12] temperate can be viewed as a control parameter regulating access to intra-packet resources, which equates temperature inverse to a cost, in entropy, of the free energy reward from the outside [54] received by the system as a result of the packet's deployment). The subjective experience of local temperature corresponds, roughly, to a level of arousal associated with object α_m . As a result, circumstances are possible when packets having low evidential support (low cumulative strength of internal associations) remain stable, separate from other packets and inaccessible to coordination with them.

2. Variations in the mode of energy delivery (level of arousal, sustained and focused attention vs. wandering and diffuse attention) cause deformations in the landscape and enable overcoming energy barriers. **Figure 13** illustrates these notions.

Maintaining focused attention underlies the experience of cognitive effort that accompanies recall or attempts to ascertain connections between some entities (e.g., objects represented by packets X_i and X_k). The experience was best described in [56], as shown in **Figure 14**.

3. Mental models are synergistic neuronal complexes that comprise packets, regulatory neuronal structures that coordinate rotation of packet vectors, and

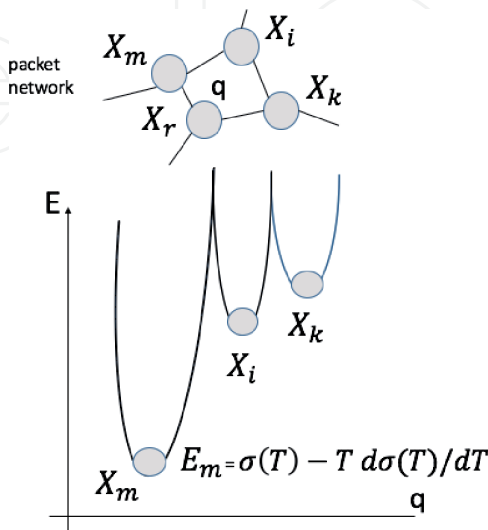


Figure 12. Here, q denotes a coordinate in the packet network space packets X_m and X_i are adjacent in the network but are not mutually accessible due to a high energy barrier that separates them. By contrast, packets X_i and X_k are mutually accessible (think of a terrain where X_i and X_k settlements are located in the same valley and are separated by a steep hill from X_m).

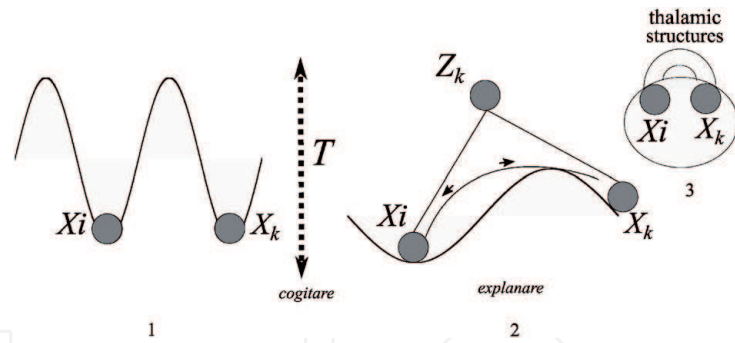


Figure 13.
1) elevated arousal combined with diffuse attention equate to increasing temperature across patches in the packet network, causing temporary lowering of energy barriers and enabling inter-packet coordination (term ‘cognition’ derives from the Latin cogitare: Shaking together [55]. 2) sustained, focused attention equate to targeted energy delivery sufficient for local lowering and overcoming of the energy barriers, enabling coordination (term explanation derived from the Latin explanare: Flatten, make level or plane (Harper-Collins Dictionary of Philosophy, 1992). 3) inter-packet coordination can involve structures residing outside packet network (i.e., cortico-thalamo–cortical connections, vs. cortico-cortical connections).

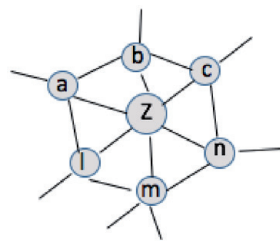


Figure 14.
The experience of mental effort. “Call the forgotten thing Z, the first facts with which we felt it was related to a, b, and c, and the details finally operative in calling it up l, m, and n. The activity in Z will at first be a mere tension; but as the activities in a, b, and c little by little irradiate into l, m, and n ... their combined irradiations upon Z succeed in helping the tension there to overcome the resistance, and in rousing Z to full activity. Through hovering of the attention in the neighborhood of the desired object, the accumulation of associates becomes so great that the combined tensions of their neural processes break through the bar, and the nervous wave pours into the tract, which has so long been awaiting its advent” ([56] p. 586).

excitatory-inhibitory connections between the packets serving to constrain vector rotation. **Figure 15** illustrates these notions.

4. Mental modeling entered the stage (i.e., Sapience emerged) when mental processes became decoupled from the motor-sensory feedback. The hypothesis is that neuronal machinery of sensory-motor coordination richly developed in the protohuman was adopted for the task of mental coordination not accompanied by any overt activities [28]. As a result, neuronal mechanisms could retain a rich repertoire of coordination capabilities but became unencumbered by the spatio-temporal constraints facing sensory-motor acts (e.g., when raising a hand, one cannot skip over intermediate positions or exceed the range and speed limits afforded by the muscular-skeletal system. By contrast, envisioning the same act does not face such restrictions).

Decoupling from motor-sensory feedback created a gateway into mental universe populated by products of composition (imagination). To yield adaptive benefits, regulatory mechanisms were needed that would curtail superfluous compositions and facilitate those that could be mapped back onto and benefit overt behavior (i.e., allow predictions). Understanding is such a mechanism: although being rooted in sensory-motor coordination, understanding allows predictions unrestricted by spatio-temporal limitations of sensory-motor processes

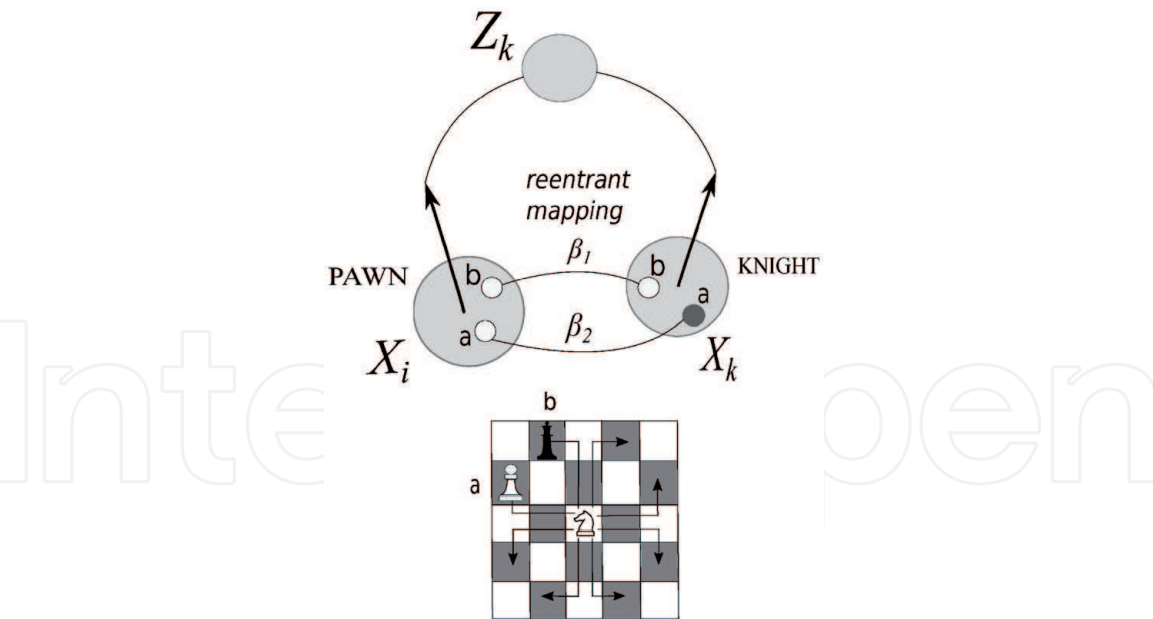


Figure 15. Understanding chess positions. White knight can move to 8 squares, thinking of possible moves involves consecutive activation of one place neuron and inhibiting the other seven in the knight packet. Place neuron responding to square *a* in the white pawn packet inhibits the corresponding neuron in the knight packet. As a result, the idea of moving knight to square *a* does not come to mind. Place neuron responding to square *b* in the knight packet excites place neuron *b* in the pawn packet, and vice versa. As a result, the idea of taking the black pawn by either the white pawn or the white knight presents itself prominently (one ‘sees’ the opportunity).

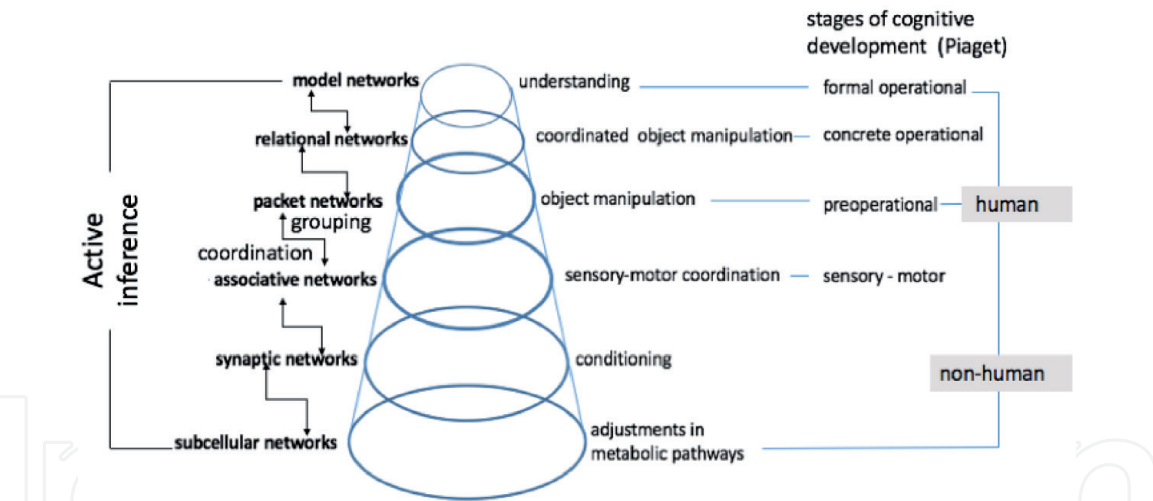


Figure 16. Functional architecture underlying active inference. The architecture comprises 6 levels, from subcellular to model networks. Subcellular networks at the bottom coordinate movement of mitochondria and substances across cell populations and inside cells. The model network on top comprises a multitude of mental models spreading across different tasks and domains. Interactions between levels are two directional: Intra-level processes form groups of elements that are treated as (composite) elements in the next level above; in turn, upper level-processes influence conditions and coordinate groupings in the level below. The packet network plays a pivotal role in the architecture, bridging levels shared by all species and those that are unique to the humans and become operational gradually in the course of an individual’s cognitive development.

or the speed of neuronal signaling. At the same time, mental models are subject to constraints of a different kind, including the explainability requirement and, crucially, limitations imposed by processes (reentrant mapping) that are inherent in the coordination mechanisms and allow eliminating superfluous degrees of freedom in the model constituents. **Figure 16** summarizes assumptions and suggestions in this part, presenting a sketch of functional hierarchy underlying active inference.

Emergence of packets underlies perception, i.e., extraction of quasi-stable, bounded feature groupings (objects) from the sensory stream (e.g., one can discern and subsequently recognize different chess pieces). The relational level is split in two – behavioral and relational proper. In the former, different behavior patterns are attributed to the objects (e.g., admissible moves are defined for knight, as in **Figure 15**). In the latter, inter-object relations get decoupled from the objects' sensory contents (e.g., coordinations in **Figure 15** make no account of the shape, color, weight, etc. of the participating pieces). Finally, operations in the model network support mental experiments (gedanken experiments) – a form of active inference most distant from the control of motor-sensory feedback. Mental experiments can entail physical experiments but do not rely on them in assessing the validity of their conclusions.

Ideas and suggestions in this section do not answer questions **a** and **b** posed at the end of Section 2 but, arguably, indicate directions for further inquiry. Question **c** will be addressed briefly in the discussion. The ideas are speculative, the next section references findings and theories in the literature that seem to agree with the ideas and might help assessing their biological plausibility.

4.3 Assessing plausibility

A thumbnail summary of the preceding two sections: Cognitive processes yield adaptive behavior via two regulatory loops: the first loop optimizes (coordinates) deployment of sensory-motor resources while the second loop coordinates deployment of neuronal resources. The first loop produces associative networks that give rise to packet networks, the second loop combines packets into nested coordinated structures (mental models). The second loop was decoupled from the motor-sensory feedback, which created an opportunity for constructing unlimited multitudes of mental models. Realization of that opportunity was predicated on satisfying two constraints: a) using a limited number of neurons and b) maintaining energy consumption below some physiologically attainable thresholds. It can be shown that mechanisms of packets and packet coordination are deployment heuristics serving to satisfy the constraints [11, 12]. Packet coordination underlies understanding, which is a form of active inference unique to Sapience. The remainder of this section references findings supporting key notions in this proposal.

4.3.1 Tuning neuronal resources

Dynamic allocation of neuronal resources implies that neurons have a degree of plasticity, i.e. their receptive fields (RF) can be changed by both the stimuli and, crucially, brain systems that regulate allocation. A body of findings in [57–62] provide ample evidence of such plasticity, including stimulus-driven adaptive plasticity, rapid attention-driven plasticity, and consolidated learning-induced plasticity. Rapid attention-driven plasticity manifests in attentional modulation of neuronal processes and underlies the ability of the brain to make coordinated changes in stimuli-driven and self-directed neuronal activities as the context and task demands change. “These transformations occur at the level of synapses, single-neuron RFs, and also at the level of brain networks” ([63] p. 252).

4.3.2 Optimizing deployment of neuronal resources

The idea to characterize cognitive processes as resource optimization has been explored repeatedly in several forms, as optimization of energetic resources [64], optimization of computing resources [65], optimization of cognitive resources [66]. The present theory characterizes cognition as deployment of neuronal resources optimized

for energy efficiency, under an exceedingly simple model (“neurons fire at stimuli”): successful allocation of neurons to streaming stimuli procures energy deposits from the stimuli and incurs energy costs (recruiting, firing, maintaining neurons), neuronal system seeks to maximize the former while minimizing the latter [12]. It can be shown that, under this model, elements of functional architecture in **Figure 16** represent heuristics delivering progressively improving energy inflows while reducing energy costs (optimal maneuvering of neuronal resources to maximize gains and minimize losses). Other major phenomena can be mapped straightforwardly onto the model, e.g. in the context of resource optimization, the short term memory/long term memory partitioning turns out to be a powerful heuristic involving breaking large optimization problems into successions of small ones thus cutting down the amount of computation while keeping the outcome in the vicinity of global optimum. Optimal allocation strategies include prediction and anticipatory recruitment (active inference), combining those with cost minimization enabled expansion and diversification of inference domains. Dynamic resource optimization requires unencumbered access to all resources in the resource pool and flexible switching between resource groupings. These notions resonate with proposals in the literature, some examples follow.

A model in [67] postulates a global workspace composed of distributed and heavily interconnected neurons, and a set of specialized modules conducting perceptual, motor, evaluative, and attentional operations. Workspace (regulatory) neurons are mobilized in effortful tasks and selectively mobilize or suppress, through descending connections, the contribution of specific processor neurons. When workspace neurons become spontaneously co-activated, they form spatio-temporal patterns that are subject to modulation by vigilance signals.

The idea of cost-reward tradeoffs is consistent with the findings in [68]. This study examined neuronal substrate responsible for balancing expected performance rewards and their cognitive costs. Single-unit recordings in monkeys provided evidence that neurons in Medial Frontal Cortex (MFC) encode associations between action sets and their rewarding values and are involved in the cost-reward tradeoffs. MFC evaluates the costs incurred in executing cognitively demanding tasks and the expected gains, and recruits control resources in the Lateral Prefrontal Cortex (LPC) as necessary for compensating performance costs. MFC responses also reflect intrinsic MFC processes inhibiting inappropriate behaviors and energizing the LPC resources involved in selecting alternative behaviors according to the rewards and penalties at stake. The ideas concerning the cost-reward tradeoffs are consistent with those in [69].

The overall notion of dynamically optimized recruitment of neuronal resources is consistent with findings in [70] associating competent performance across multiple domains (“general intelligence”) with selective recruitment of lateral frontal cortex in one or both hemispheres. These same frontal regions were found to be recruited by a broad range of cognitive demands, thus suggesting that “general intelligence” derives from flexibly switching recruitment between different neuronal groups. Another facet of neuronal processes implicit in the idea of neuronal resource optimization is “neuronal reuse”, i.e. engaging the same circuitry for different behavioral purposes [71]. Combining quasi-stable neuronal packets without changing the packets or the underlying mosaic of associative links is a form of reuse. Improving energy efficiency can be a factor in the optimization of cerebral cortex layout and physical embedding of processing networks in the brain volume [72]): minimization of total connection length [73] reduces energy costs of signal propagation.

4.3.3 Improving energy efficiency

Neuronal processes consume significant amount of energy, consumption increases with activity which demands local and global changes in metabolic rates

and blood flow. Mechanisms of efficiency and energy transduction in the brain have been investigated in numerous studies [74–78]. Energy is produced through oxygen consumption mediated by the mitochondrial respiratory chain generating the high-energy phosphorous metabolite (adenosine triphosphate, or ATP). The carbon source that supports the oxidative metabolism is predominantly glucose. About 20% of the total oxygen consumption in the body takes place in the brain. A detailed account of energy consumption was obtained in a recent study utilizing ^{31}P -MRS in vivo imaging of the human brain [79]. It was determined that approximately 5.7 kg of ATP molecules is produced and utilized by the cortical gray and white matter in a day, which is equivalent to the complete oxidative combustion of 56 g glucose per day and is almost five times the total weight of the gray and white matter (≈ 1.2 kg). The energy expenditure of a single cortical neuron is 4.7 billion ATPs per second (compared than 3.3 billion ATPs/neuron/sec estimated for the rat brain). Approximately 67–75% of the total energy expenditures is used for neurotransmitter signaling and electrophysiological activities involved in sustaining neuronal functions [79].

It has been long recognized that the high energetic cost of human brain function, which is 10 times higher than what would be expected from its weight alone, can only be maintained through efficient energy use [80, 81]. Accordingly, theories were advanced suggesting that brains evolved to be metabolically efficient [82–84] which implies that representations of events and actions should be sculpted to involve as few action potentials and active synapses as possible. For optimum efficiency, less than 4% of a population of cortical neurons should be activated to represent a new event. Neural mechanisms associated with attention restrict the volume of cortex in which activity is elevated [85]. The arrangements of neuronal systems are thought to allow maximum communication speed with minimal energy expenditures [86].

Massive data was accumulated demonstrating reduction of metabolic costs in the organization of motor performance and regulation of movement economy [87–93]. As noted in [94], metabolic determinants of physical action organization might not be the same as those determining cognitive action organization. However, it stands to reason to assume that the principle of cost minimization applies in both domains.

Analysis in [85] concludes that strategies directed at maximizing metabolic efficiency are indeed used by the brain. In particular, a) fine axon collaterals reduce the number of ions required to transmit an action potential, by reducing membrane area, b) the arrangement of neurons in maps reduces the distance the potentials must travel and c) sparse codes reduce the number of action potentials required to represent events. **Figure 17** indicates that suggestions in the present theory resonate with those formulated in [85] and other studies.

4.3.4 Neuronal packets are building blocks in cognitive processes

The idea that dynamically formed neuronal groupings (assemblies, ensembles) are the basic functional units in neuronal processes was advanced by [95] and subsequently developed in the Theory of Neuronal Group Selection (TNGS) by Gerald Edelman [96–98] and explored in other studies. For example, [99] suggests that acquisition of motor skills involves development of motor primitives amenable to adaptive re-combination (arguably, motor primitives are rooted in the underlying neuronal assemblies), [100] conceptualizes mental synthesis as a synchronization of independent neuronal ensembles, etc. Hebb's idea received experimental support in a number of recent findings: studies in [86, 101] demonstrated existence of neuronal assemblies entering into different combinations as the tasks and conditions change. Assemblies observed in [86] comprise a few dozen neurons each and can

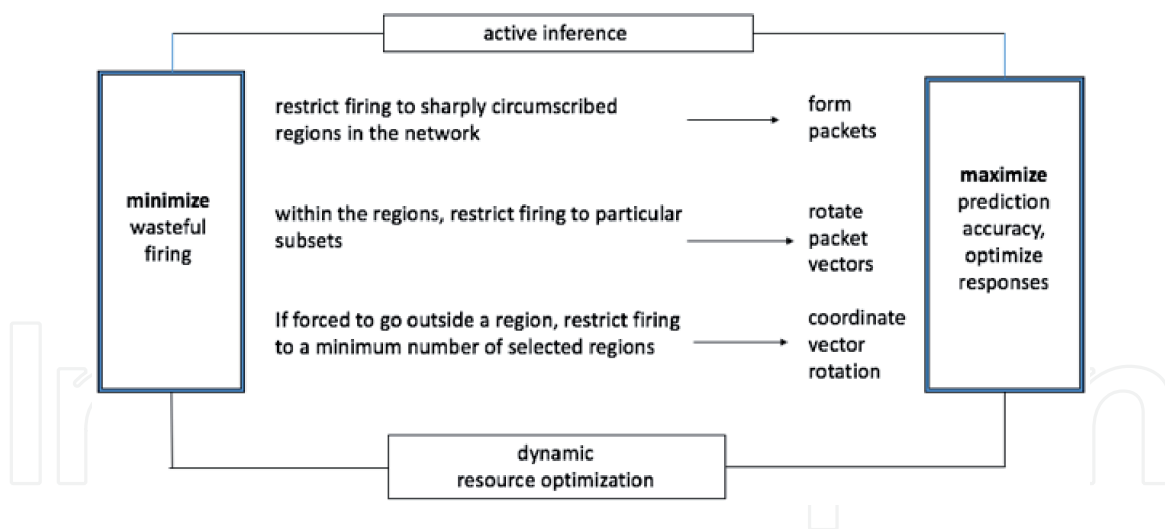


Figure 17.

Connected associative network allows unrestricted signal propagation, i.e., excitation of any neuron can ignite excitation spreading that will, eventually, engulf the entire network. Formation of packets and operations on them minimize spreading, confining excitation to the smallest subset of neurons producing the largest expected energy gain. Dynamic resource optimization boils down to suppressing wasteful firing and facilitating beneficial firing, i.e. yielding maximum prediction accuracy and response composition optimal under the prediction. In that sense, resource optimization is an engine of active inference.

be interlaced within the same volume. It was suggested that “elementary neuronal groups are prescribed Lego-like building blocks of perception and that acquired memory relies on combining these elementary assemblies into higher-order constructs” [86]. Both studies suggest that their findings reveal a synaptic organizing principle (i.e., grouping) that is common across animals.

An important elaboration of the notion of assembly received in the idea of *synergistic structural units* formulated in [102–106]. Synergistic structural units can be combined into task-specific groupings and, crucially, are amenable to “nonindividualized control”, that is, their constituent elements can be controlled by a few task-related variables (goals) [103].

The notion of ‘neuronal packets’ builds on the idea of Hebbian assembly and is consistent with the finding and suggestion referenced above. However, the notion offers two crucial extensions to the idea, as follows: a) neuronal packets form as a result of phase transition in associative networks causing some subnets to fold into cohesive units (packets) and b) folding establishes energy barriers at the packet boundaries. Stated differently, boundary energy barriers implement Markov blankets separating packet internals from the surrounding network [30]. More precisely, the height of energy barriers equals free energy per unit of surface area (surface tension) determined by the total membrane surface in the packet’s boundary neurons (i.e., packet’s Markov blanket, see **Figure 9**). Analysis in [85] identified reduction of membrane areas in individual neurons as a factor contributing into brain’s metabolic efficiency. In a similar way, thermodynamically-driven tendency to minimize packet surface areas [27] contributes to the metabolic efficiency of neuronal processes (see **Figure 17**).

4.3.5 Coordinated rotation of packet vectors

Phase transitions transform groups of associated neurons into cohesive functional units amenable to synergistic control and re-combination with other units (reuse). The present theory defines coordinated rotation of packet vectors as a form of synergistic control, extending control mechanism described in [51–53, 107–111] from controlling overt movements to controlling mental ‘movements’ (i.e., packet vector rotation and coordination, see **Figure 15**). This generalization is consistent

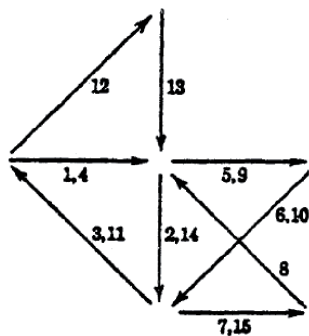


Figure 18. Neuronal assemblies were conceptualized as complex structures affording different trajectories for excitation propagation (adopted from [95]).

with the original concept of neuronal assemblies in [95] envisioning the possibility of the assemblies producing different responses constituted by different excitation trajectories within the assembly, as shown in **Figure 18**.

The notion of packet vector trajectories appears to be consistent with the findings in [112] demonstrating that memorization involves formation of specific sequences of spike bursts in the cortex that are replayed during retrieval. The function of coordination (neurons Z_k in **Figure 15**) can be carried out by components of basal ganglia, thalamus and other structures. In particular, [113] suggests that basal ganglia chunks the representations of motor and cognitive action sequences so that they can be implemented as performance units. Studies in [114] uncovered activities in basal ganglia circuits that encoded sequences as single actions. Besides start/stop signaling and sequence parsing, these neurons displayed inhibited or sustained activity throughout the execution of the sequences. This sustained activity co-varied with the rate of execution of individual sequence elements, consistent with motor concatenation. Direct and indirect pathways of basal ganglia were concomitantly active during sequence initiation, but behaved differently during performance. Thalamic relays also play a critical role in coordination [115, 116]. The cerebellum is also involved in the detection and generation of sequences [117].

Cortical coordination and dynamics have been analyzed in [118–121] concluding that “the formation of neural context through the coordinated mutual constraint of multiple interacting cortical areas, is considered as a guiding principle underlying all cognitive functions” ([120] p. 140). The present theory agrees with that conclusion and suggests neuronal mechanisms instantiating the idea. In particular, the theory defines mental models as tightly coordinated gestalts, or *structural units* where changes in one component cause reciprocal changes in the other ones (e.g. when one hand is used to lift a heavy object from a tray supported by the other hand, increasing effort in one hand is concomitant with relaxation in the other one – hands form a structural unit (Gelfand et al). The same coordination mechanism underlies operation of mental models, e. g. in a catapult model, increasing distance to the target entails the realization that projectile need to be shifted in the opposite direction). The modeling mechanisms includes coordinated vector rotation and reentrant mapping.

4.3.6 Reentrant mapping

The hypothesis that reentrant signaling serves as a general mechanism to facilitate the coordination of neuronal firing in anatomically and functionally segregated cortical areas and in the thalamus is one of the main tenets in the Theory of Neuronal Groups Selection (TNGS) [122–124]. According to TNGS, neurons

belonging to different cortical areas are reciprocally interconnected by reentrant networks of excitatory axons, and each cortical area is also reentrantly interconnected by large numbers of axons to one or more nuclei of the thalamus. These thalamocortical and cortico-thalamic reentrant connections modulate brain arousal and help determining which of the patterns of environmental signals arriving in the thalamus from the environment will be relayed on to the cortex. They also participate in the execution of timed, sequential, or willed processes, such as manipulating mental constructs, or issuing segmented motor commands [124]. The present theory is consistent with TNGS principles, making reentrant mapping (or bidirectional coupling [113]) integral to the mechanisms of modeling and understanding (see **Figure 15**).

4.3.7 Energy landscapes – a missing link in cognitive neuroscience

It has been long recognized that the concept on neuronal assembly leaves the issues of stability and borders undetermined (how does the brain ‘know’ where one assembly ends and another begins, how does a neuron ‘know’ to which assembly it belongs, what keeps neurons in an assembly together, etc.)? In the original conceptualization [95], waves of excitations develop and reverberate inside assemblies - this notion indicates intuition of assembly borders but that intuition was not made explicit. The original conceptualization in [95] entailed a possibility that activity in any assembly will spread to other assemblies and ultimately to the entire cortex or even the total brain, resulting in pathological overactivity, as in seizures. To cope with the problem, the idea of a “threshold control mechanism” was introduced [125] with the subsequent elaborations placing the mechanism in the basal ganglia or the hippocampus. The idea was that a cell assembly “holds” at a threshold θ when at that threshold all the neurons of the assembly, once excited, stay active due to their reciprocal excitatory connections. Manipulation of the thresholds was envisioned as follows (compare to **Figure 13(1)**).

“A periodic operation (colloquially called the “pump of thoughts”) may involve the following steps. Given a certain input I, the threshold is lowered so that the set of active neurons FI will go over into a larger set F'I. This will encourage the ignition of cell assemblies. As the threshold is again raised, activity is smothered and only the most strongly connected cell assembly will survive. A new cycle beginning again with a lowered threshold will bring in new cell assemblies. They may include an even more strongly connected cell assembly, which will be the next one to survive when the threshold is raised. The evolution will be in the direction of the most strongly connected cell assemblies.... One may express this by saying that the system hunts for an interpretation of the input, or that it ‘thinks’ ([125] p.177).

Independently from the proposal in [125], the idea of threshold regulation was advanced in a theory of movement coordination (λ theory) in [126–128]. According to λ theory, coordination of motor actions involves centrally controlled resetting of the threshold positions of body segments. Deviations from the threshold positions (e.g., restive muscle length) trigger resistive forces, detection of differences between the centrally set threshold positions and the sensory-signaled actual positions cause activation of neuromuscular elements seeking to diminish the difference. The crucial assumption is that thresholds are changed by descending fibers that influence membrane potentials of motoneurons in motor cortices, either directly or via interneurons [126].

Arguably, theories of threshold regulation [125, 126] are motivated by intuition similar to that expressed in **Figures 13-15**. In the present theory, boundary barriers

are an intrinsic property of neuronal assemblies (packets), regulation of barrier height involves changes in membrane potential in neurons residing in the packet's MB. Boundary energy barriers make assemblies distinct, quasi-stable and immersed in energy landscapes. The landscape curtails activation spreading, by imposing energy costs on inter-assembly transitions.

More precisely, the present theory postulates that boundary barriers establish energy landscapes across packet networks [10, 12]. Accordingly, formation of packets can be viewed as a form of folding, analogous to the folding of proteins and other complex molecular structures [129–131]. As in proteins, the folding of packets is a spontaneous process obtaining stable (equilibrium) configurations of minimal free energy [27]. Stability is maintained within some ranges of temperature variation (packets dissolve when $\sigma(T) \approx T d\sigma(T)/dT$, $E_m \rightarrow 0$, the constituent neurons become absorbed into the surrounding packets). Within the multidimensional energy surface, packets' Markov Blankets and the corresponding cutsets (links connecting MB to the surrounding packets) form attraction basins in the neighborhood of local minima, connected by saddle points. As a result, attentive navigation of the landscape involves energy-demanding (effortful) basin-to-basin transitions. Deformations in the energy landscape determine changes in the accessibility of neuronal packets. Presumably, transitions are controlled by frontal /prefrontal networks and thalamic structures.

A number of experimental results appear to agree with the proposal. Findings in [132] demonstrating fast transitions between separated states of cortical activity involving distinct neuronal groups appear to agree with above proposal. Findings in [133] indicate that thalamic cells respond selectively to complex percepts and concepts conferred on them by the cortical assemblies in whose activation they participate. The cortico-thalamo-cortical pathways provide connections between different cortical loci which have higher reliability than the direct cortico-cortical routes, and play crucial role in orchestrating activation of those assemblies). Important findings in [134] demonstrated that brain network are structured in a manner optimized for network control, which includes increased controllability and reduced synchronizability (controllability characterizes the ease of switching from one dynamical state to another, traversing energy landscape (see Figure 130; synchronizability characterizes the ability for regions in the network to support the same temporal dynamical patterns).

The idea of energy landscapes in brain systems remained purely speculative until the recent pioneering studies in [135–137] applied modern analytic and modeling techniques (e.g. network disconnectivity analysis) to fMRI data, seeking to define energy landscapes in Default Mode Network (DMN) and Fronto Parietal Network (FPN). It was determined that DMN energy landscape consisted of two groups of low-energy local minima that are separated by a relatively high energy barrier. Within each group, the activity patterns of the local minima were similar, and different minima were connected by relatively low energy barriers. In the FPN, all dominant local minima were separated by relatively low energy barriers such that they formed a single coarse-grained global minimum. The height of energy barriers separating local minima influences the rate of inter-state transitions. Accordingly, transitions in DMN occur at a low rate while transitions between local minima in FPN occur more easily. The notion that brain operates at the edge of instability and transits between low energy states has been explored in multiple studies [50, 138]). It appears that the notion of brain energy landscapes was introduced in [10, 12], and experimental mapping of energy landscapes was attempted for the first time in [136].

To summarize, the folding of subnets in associative networks forms packets separated by Markov Blankets from the rest of the network. Packet Markov Blankets are constituted by boundary energy barriers that make packets distinct, quasi-stable

stable (i.e., amenable to modification but at substantial energy cost) and synergistic (i.e. amenable to control by a few variables and coordination with other packets). Boundary barriers establishes energy landscapes across packet networks and determine both kinematic (inter-packet transitions) and dynamic properties of neuronal organization.

4.3.8 Accommodations

It was suggested that lateral inhibition prevents neuronal assemblies from encroaching on each other while the tendency towards reducing surface tension in the packets favors their coalescence (minimizing the amount of free energy in the surface). Arguably, the interplay of the opposite tendencies drives ‘accommodation’, that is, spontaneous adjustments inside the neuronal systems following changes resulting from interactions with the environment [27].

The notions of assimilation, accommodation and cognitive equilibration were introduced in [18] denoting, correspondingly, integration of new information into the existing structures, re-organization of those structures, until a state of equilibrium is reached obtaining a sufficient degree of integration via a minimal amount of structural changes. According to the present theory, assimilation involves changes in the distribution of synaptic weights, that trigger waves of packet re-structuring propagating throughout the packet network (the accommodation). In this way, the requirement of spontaneous re-structuring is inherent in the notion of neuronal packets immersed in energy landscapes. On that view, the overall functional architecture of the cognitive systems was reduced to three modules: associative cortices, reticular formation controlling arousal level, and a frontal/prefrontal module controlling landscape navigation. Accommodation and assimilation are confined to packet networks [10].

Recent experimental findings and theoretical proposals [139] envision functional architecture comprising Default Mode Network (DMN) [140], Salience Network (SN) [141] and Task Control Network (TCN) [142], as follows. A DMN is a large network comprises hubs in medial prefrontal cortex, posterior cingulate/precuneus and angular gyrus becomes active under conditions of wakeful rest, i.e. when person is not engaged in any task. The SN comprises a suite of brain regions whose cortical hubs are the anterior cingulate and ventral anterior insular cortices while the TCN (a cingulo-opercular task-control network) is anchored in the dorsal anterior insula and the frontal operculum. SN detects behaviorally relevant stimuli and recruits neural resources to orchestrate responses. For the latter, the SN engages the TCN (or Central Executive) whose functions include maintaining relevant task set or orchestrate switching to a new task set in response to shifts in the salience landscape.

Significantly, a comprehensive study in [143] compared functional networks in the brain during task performance (active brain) and at rest (resting brain), concluding that the full repertoire of functional networks utilized in active brain (Active Brain Networks, or ABN) remains continuously active in the resting brain (Resting State Networks, or RSN, including the “default mode network”). The study applied independent component analysis (ICA) and other modern techniques to two sets of fMRI functional imaging data: “active brain” data in the BrainMap data base collected from over 30,000 subjects, and resting brain data collected from 36 subjects. The ICA decomposition was conducted at two resolution levels, 20-component analysis and 70-component analysis, with the higher resolution analysis revealing subnetworks in the primary networks determined at the lower resolution level. It was found that primary networks split into subnetworks in both active and resting data in almost identical ways, maintaining greater functional

(temporal) correlation between subnetworks within a primary network than across primary networks. Analysis in both levels produced converging results: close to 70% overlap in the composition of Active State Network and Resting State Networks. The analysis concludes with an admission: “Although we have shown that activation networks are mirrored in resting data, we must acknowledge that this does not begin to answer the question of why the brain’s many regions continue to “function” (with large amplitude fluctuations) when the subject is at rest, and even when the subject is asleep and under anesthesia” [143].

It appears that these findings are consistent with the proposal in [10] envisioning waves of accommodating adjustments in packet networks. Moreover, the adjustment requirements are inherent in the notion of packet networks. In particular, the hypothesis is that variations in temperature and synaptic weight distribution across packet networks cause changes in the resting membrane potentials [144] in the MB neurons, thus creating potential gradients in the packet network causing adjustments in the energy landscape and re-distribution of neurons seeking packet configurations in the vicinity of global energy minima. Stated differently, energy landscape is “frustrated” [131] due to conflicting tendencies of lateral inhibition and lateral coalescence. Spontaneous re-organizations in packet networks to resolve frustration move the system in the direction of cognitive equilibrium. Possibly, neuronal avalanches are a form of such re-organization, playing a role in maintaining network stability and preventing runaway excitation [145]. **Figure 19** makes suggestions regarding the placement of packet networks in the tri-partite architecture [139, 146].

Figure 19 suggests that DMN/SN/CEN interplay focuses on the engagement of prefrontal areas in coordination activities, i.e., formation of relations and operations on relational networks. Accordingly, it can be expected that prefrontal damages are likely to cause severe deficits in integration of relations. The order of operation in the DMN/SN/CEN system is, roughly, as follows: a) the Central Executive Network includes the agency of attention and controls attention focusing and other processes engaged in the performance of cognitive tasks, b) the Default Mode Network becomes active when the person remains awake but no tasks are pursued, c) the Saliency Network administers switching between CEN and DMN.

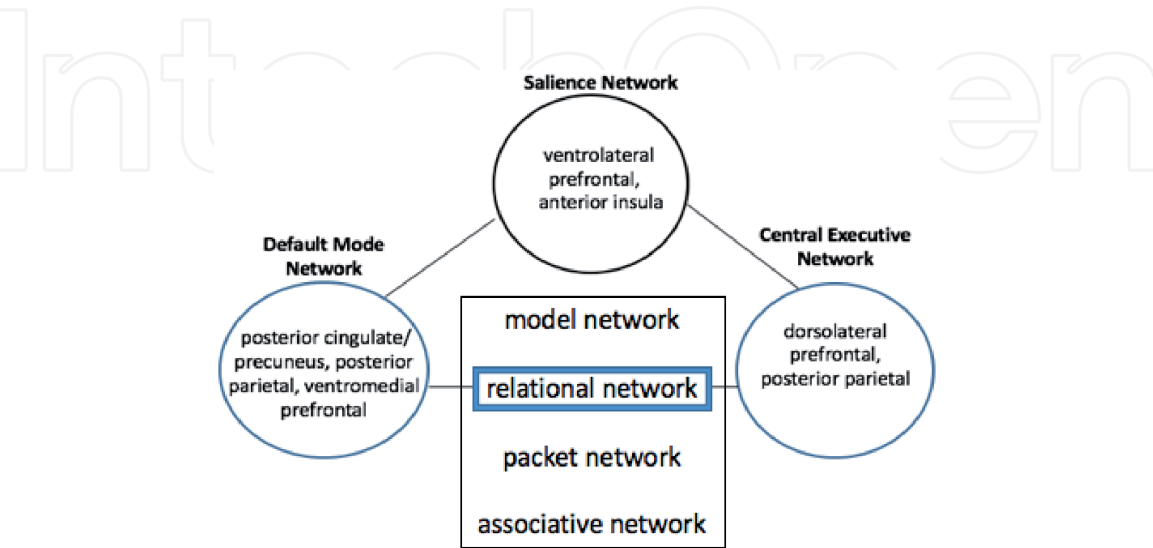


Figure 19. Operations on networks underlying understanding capacity involve an interplay between default mode network and central executive network (CEN). Saliency network coordinates switching between DMN and CEN [146]. This general architecture allows more detailed mapping onto anatomical structures in the brain underlying functional organization [147–149].

4.3.9 Decoupling

The present theory attributes emergence of human understanding to evolutionary developments causing decoupling of regulatory processes from the sensory-motor feedback loops [10, 28]. The idea is consistent with suggestions in [150] regarding evolutionary origins of human cognition. Analysis [150] focuses on the development of cerebral cortex, pointing at its vast expansion in the humans relative to other primates (the cerebral surface area is 120 cm² in the macaque and 960 cm² in the human) and disproportionate expansion of distributed association regions within the cortex. The hypothesis is that rapid expansion of the cortical mantle may have decoupled (“untethered”) large portions of the cortex from sensory hierarchies and resulted in the development of networks that either control processes in the sensory networks or are engaged in parallel activities that are “detached from sensory perception and motor actions – what one might term ‘internal mentation’” [150].

5. Discussion and conclusions

In a letter to Nature Neuroscience entitled “What does ‘understanding’ mean”?, the author confesses that “upon reflection, it is depressing, if not scandalous, to realize how rarely I ask myself this” [151]. Arguably, the letter’s intent was not to confess ignorance or lack of interest but to point out that a critically important issue has been long neglected. There is nothing that one is familiar more intimately and directly with than sensations of confusion, mental effort and understanding (except, perhaps, for the sensations of one’s own breathing and heart beating), yet the issue of understanding has not been receiving significant attention in cognitive sciences (see some discussion in [27, 152]). The intent of this chapter was to suggest that a theory of understanding might be within reach (and grasp), requiring the synthesis of new ideas and the long existing ones, re-evaluated in the light of new data. The proposed theoretical framework is that of active inference [13, 14] carried out under the requirements of limited neuronal pool size and minimized energy expenditures. Within that framework, the meaning of ‘understanding’ reduces to optimization strategy in the deployment of neuronal resources that enables expanding domains of inference while minimizing expansion costs. In subjective experiences, the meaning of understanding reduces to attaining ‘grasp’, i.e. unifying some disparate entities, in a coordinated relational structure that enables relational [153, 154] and other forms of reasoning. Attaining ‘grasp’ can be accompanied by cognitive strain and culminates in exhilaration and euphoria making the activity self-rewarding (the Greek *euporia* stands for ‘easy passage or travel’ while its opposite *aporia* denotes ‘difficulty or impossibility of passage’ [48]). This section will compare the present proposal to some findings in the literature, aiming to suggest directions for further research.

5.1 Mental simulations

The phenomenon of mental modeling (mental simulations) has been addressed in a number of studies [155–157], focusing on the “paradox” of endogenously-driven mental activity:” how can findings that carry conviction result from a new experiment conducted entirely within the head” [155]. Data has been accumulated demonstrating that mental simulations engage mechanisms that are different from those involved in reasoning based on descriptive knowledge, exhibit analogue properties, and can produce correct inferences when descriptive knowledge is lacking. At the same time, it was observed that mental simulations proceed in a piecemeal fashion (not a holistic image) [157].

The present proposal pivots on the notion that mental modeling was made possible by decoupling regulatory processes from the motor-sensory feedback, which shifted the power of conviction from experiments in the world to experiments in the head (e.g., arguments in Pythagorean theorem are entirely convincing but not amenable to experimental verification). On the account of the present theory, the experience of understanding accompanies formation of tightly coordinated gestalts which, simultaneously, afford some degrees of freedom to their constituents. Exploring these degrees of freedom can indeed proceed in a step-by-step fashion, i.e. experimental findings in [157] and are not incompatible with the theory. Other proposals in the recent literature addressing the role of mental simulation [158] resonate with the key notions in this theory.

5.2 Transient assemblies and the searchlight hypothesis

The operation of focused attention was compared to a searchlight that shifts between and thus helps forming conjunctions of separate attributes or features of perceived objects [159]. It was further proposed that functions of the “searchlight” are carried out by activity bursts in thalamic nuclei while conjunctions are implemented by rapidly modifiable synapses (called Malsburg synapses), orchestrated by the bursts to produce transient cell assemblies [160].

Notwithstanding suggestions in [159, 160] concerning transient assemblies, considering the role of focused attention in manipulating quasi-stable assemblies (packets) calls for a different metaphor. A neuronal packet is a superposition of multiple behavior patterns afforded by an object. Overcoming energy barrier and shifting attention from outside the packet to the inside (see **Figure 13(2)**) actualizes one of the patterns. Think of ‘grasp’ as seizing an object and holding it in a closed fist, followed by opening the fist and holding the object in an open palm. With the eyes closed, one needs to run the fingertip of another hand over the object in order to discern its shape. The point is that concentrating attention amounts to focusing energy delivery on particular neurons causing their excitation or inhibition, which gives rise to the experience of a behaving object. In short, both the searchlight and the fingertip metaphors define attention as physical actions applied to neurons. However, the former metaphor conjures up an image of a wandering light beam falling on the elements of neuronal structures and thus making them discernable to the “mind’s eye” while the latter one connotes the image of a finger (or stick) ‘tapping’ on the neurons, which seems to better represent the notion of physical action.

5.3 Understanding and language

The discovery of mirror neurons inspired hopes that understanding of the origins of language can be “within our grasp” [161]. Mirror neurons discharge during active movements of the hand or mouth (or both) performed by the subject or observed being performed by others (hence, the mirror neurons). It was hypothesized that the latter feature establishes a bridge from ‘doing’ to ‘communicating’, or from acting to message sending [161, 162]. Other hypothesis concerning language origins attribute its emergence to internal, as opposed to communicative, functions [35–37] and conceptualize language mechanisms as the manipulation of neuronal assemblies [163, 164]. This theory offers an opinion that seems to unify all three hypotheses, as follows.

First, note that mirror neurons were determined to be of three types: ‘grasping with the hand’ neurons, ‘holding’ neurons and ‘tearing’ neurons [161]. Apply these notions to manipulation of mental ‘objects’ (as opposed to physical ones) and assume that ‘grasping with the hand’ denotes formation of a packet, ‘holding’ denotes the state when attention is “hovering outside” the packet (see **Figure 14**),

and ‘tearing’ denotes entering the packet and experiencing the contents. A reversible ‘holding’ – ‘tearing’ transition corresponds to set operation: a manifold of features is experienced as a unity (one object) devoid of (separated from) any sensory contents, followed by experiencing a series of sensory features comprised in the object.

Next, think of watching a play performed on the stage, and then consider the same play being read to you. In the latter case, assume that the cast of characters and all the names have been removed so only the text proper remained. It is not hard to realize that figuring out what is going on might be possible but extremely difficult, requiring forming and comparing different word combinations (e.g. “The queen, my lord, is dead. She should have died hereafter...” – who is talking here? Note that you are facing no such challenges when watching the play). Finally, imagine that only the cast of characters and names are extracted from the text and the rest is discarded. Clearly, it can be very hard but possible to make some sense of the former version while the latter one makes no sense at all. It is also evident that the range of understanding in the former version will be restricted to a few characters and a few consecutive episodes, with the text becoming an impenetrable mess after that. Restoring the original text (putting the names back where they belong) resolves the otherwise insurmountable difficulty. Here comes a tentative proposal:

Emergence of language followed decoupling from the sensory-motor feedback while retaining the mechanisms of sensory-motor coordination. Language emerged as a means to support mental coordination over an expanding variety of mental objects, by adopting the mechanisms of communicative signaling and re-purposing them for self-signaling (communicative signals make an animal aware of a predator or other condition without direct sensory confirmation of that condition). Symbols (labels) are implemented as neuronal assemblies [163, 164] or ‘symbol packets’ attached to ‘object packets’, ‘symbol packets’ have no sensory content except for the minimum required for making them distinct. Symbols make one roughly aware of the contents of a packet without the expense of entering and examining these contents, thus facilitating landscape navigation (think of labels attached to drawers that need to be pulled with effort). The process of thinking alternates reversibly between the packet arrays (roughly, between words and images and actions they signify). Understanding phrases involves syntactic coordination and, crucially, substantive, or grounded [165] coordination (i.e., between the objects and activities signified by the words). Findings in [166] demonstrating “grasping ideas with the motor system”, i.e. activation of the motor cortex by words referring to bodily actions, even idiomatically, other results [167] appear to support these contentions.

5.4 Cognitive disorders

Pathological malfunctions in the operation of the DMN/SN/CEN system (**Figure 19**) can cause breakdowns in the regulation of energy landscapes (energy barriers are rigid and remain abnormally high or abnormally low), entailing a range of cognitive disorders. In particular, abnormally high barriers hamper correlation between cortical areas and interactions between frontal and parietal, neostriatum, and thalamic areas involved in attention control, which can manifest in performance impairments characteristic of the autism spectrum disorders [168–170]. By contrast, abnormally low barriers entail destabilization and disintegration of neuronal packets, leading to irreversible memory losses and other impairments characteristic of the Alzheimer’s – type disorders (e.g., subjects can be expected to fail clock drawing tests due to the inability to recollect proper elements and/or their respective positions [171]. In general, abnormally high energy barriers degrade functional connectivity between memory elements (percepts, concepts) while abnormally low barriers degrade the elements. It appears possible to relate a variety

of cognitive disorders (e.g. different forms and stages of dementia) to persistent abnormalities in energy landscapes, which can potentially lead to new insights and unified approaches in the diagnosis and treatment.

To conclude, this chapter suggested a hand-in-glove relationship between an information-theoretic account of cognitive processes (active inference) and a thermodynamics-centered account asserting that neuronal mechanisms underlying active inference are sculpted by physical conditions in the brain limiting its volume and energy supply. Active inference has been conceptualized as a regulatory process allowing organisms to operate within the sensory-motor feedback loop. This is accomplished by forming generative models that anticipate consequences of overt actions as those are reflected in the sensory inflows, followed by adjustments that reconcile the actions and the models in a manner serving to satisfy the survival and other needs. This chapter applied the active inference framework to define regulatory mechanisms decoupled from the motor-sensory feedback loop, under the notion of energy-minimizing deployment of neuronal resources.

Advanced theoretical analysis seeking to unite conceptual foundations of the physical sciences and biology is uncovering a profound unity of the information-theoretic and thermodynamics-centered viewpoints, spanning the range from inanimate matter to the most complex life forms [172]. Moreover, recent experimental findings demonstrate the possibility of information-to-energy conversion [173]. Analysis indicates that self-organization obtains access to progressively higher degrees of order and organization in the channels of energy transduction [172]. The notion of increasing levels of coordination in the brain functional architecture, from subcellular processes to mental modeling, appears to agree with this general principle. Evolutionary climb to the upper reaches of organization manifested in creative thinking was made possible by minimizing energy costs in every step. On the present theory, active inference is the result and expression of that underlying, thermodynamically-enforced frugality.

In machine intelligence, the bulk of effort has been concentrated on learning techniques derived from the perceptron idea (conditioning). This proposal suggests advancing from machine learning to machine understanding, requiring a different conceptual foundation. It has been argued that human understanding requires awareness, and physical processes in the brain that evoke awareness might not be amenable to computational simulation [174]. Notwithstanding these arguments, it appears possible to construct artifacts possessing a level of understanding that does not reach human heights but exceeds those accessible to the conventional technology.

It feels appropriate to end this chapter by giving credit to those whose foresight brought them long ago to conclusions similar to those expressed here:

“It is worth while to speculate about cell assemblies as an alternative to feature detectors and hierarchies of classificatory units. These concepts are related to Perceptrons. Similarly, cell assemblies would find their technological analogue in a (non existing) Conceptor. ... It would be surprising if it turned out that the real brain makes use only of one or the other scheme. Most likely the two schemes are used in combination, with the hierarchical organization predominating at the sensory and motor periphery of the nervous system, and the cell assemblies in between. From this point of view the cerebral cortex would seem a good place for cell assemblies, and we have seen that it contains the necessary equipment” [125] p. 187

Acknowledgements

The author is grateful to Karl Friston, Rosalyn Moran, Thomas Parr, Maxwell Ramstead, Vlad Krasnopolsky, Todd Hylton, Mark Latash for insightful comments

and helpful discussions. Special thanks are due to Raj Malhotra of AFRL for supporting this and earlier work and for his help in evaluating ideas and steering the efforts towards their application. This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-20-1-0013 and by the Air Force Research Laboratory under award number FA8650-19-C-1692. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the United States Air Force.

IntechOpen

IntechOpen

Author details

Yan M. Yufik

Virtual Structures Research, Inc., Potomac, MD, USA

*Address all correspondence to: imc.yufik@att.net

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Craik K. J. W. *The Nature of Explanation*. Cambridge University Press, Cambridge, UK. 1943.
- [2] Rook L. L. Mental models: A robust definition. Faculty of Business - Papers (Archive). 1361 <https://ro.uow.edu.au/buspapers/1361>, 2013
- [3] Doye, J. P. K., Wales, D. J. On potential energy surfaces and relaxation to the global minimum. *J. Chem. Phys.* 1996: 105, 8428-8445
- [4] Moray N. Identifying mental models of complex human-machine systems. *Int. J Industrial Ergonomics* 1998: 22, 293-297.
- [5] Moray N. Models of models of... mental models. In: Moray N. (ed). *Ergonomics: Major Writings*. Taylor and Francis, London, UK, 2004.p. 506-526.
- [6] Rouse W. B., Morris N.N. On looking into the black box: prospects and limits in the search for mental models. *Psychological Bulletin* 1986: 100, 349-363.
- [7] Johnson-Laird P.N. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge University Press. N.Y. 1983
- [8] Johnson- Laird P. N. Mental models and deduction. *Trends in Cognitive Sciences* 2001; 5, 10, 434-44
- [9] Gentner D., Gentner. D. R. 1983. Flowing waters or teeming crowds: mental models of electricity. In: Gentner, D.A. Stevens A. (eds.) *Mental Models*. Lawrence Erlbaum, Hillsdale, New Jersey, USA, 1983; p. 99-130.
- [10] Yufik Y.M. Virtual Associative Networks: A framework for cognitive modelling. In: Pribram K.H. (ed). *Brain and Values*. Lawrence Erlbaum Associates, New York, NY, 1998a: p. 109-177
- [11] Yufik Y.M. Probabilistic Resource Allocation System with Self-Adaptive Capabilities, US Patent 5, 886, 219, 1998
- [12] Yufik, Y.M. How the mind works: An exercise in pragmatism. In *Proceedings. Int. Joint Conf. Neural Networks (IJCNN 02)*, Honolulu, HI, USA, 2002: p. 2265-2269.
- [13] Friston K. J. Life as we know it. *Journal Royal Society Interface* 2013: 10, 20130475.
- [14] Friston, K. J. The free-energy principle: A unified brain theory? *Nature Review Neuroscience* 2010; 11, 127-138
- [15] Herculano-Housel S. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers Human Neuroscience*. 2009; <https://doi.org/10.3389/neuro.09.031.2009>
- [16] Schurmann T., Grassberg T. The predictability of letters in written English. arXiv:0710.4516 [physics. soc-ph] 2007.
- [17] Chang M-C., Yang A. C-C., Stanley H.E., Peng C. K. Measuring information-base temperature and energy of literary texts. *Physica A* 2017; 468, p. 783-789
- [18] Piaget J. 1975. *The Development of Thought: Equilibration of Cognitive Structures*. The Viking Press: New York, NY. 1975
- [19] Freud, Z. *The Joke and Its Relation to the Unconscious*. Penguin Classics, 2003
- [20] Bower T.G.R. *Development in infancy*. W.H. Freeman & Co. San Francisco, Ca. 1974
- [21] Wright T. *William Harvey: A life in circulation*. Oxford University Press. 2013

- [22] Aird W. C. Discovery of the cardiovascular system: From Galen to William Harvey. *Journal of Thrombosis and Hemostasis* 2011; 9, 118-129.
- [23] Rovelli C. Reality is not what it seems. Riverhead Books. 2014
- [24] Lehrer J. How we decide. Houghton Mifflin Harcourt. 2009
- [25] Groot, A.D. Thought and Choice in Chess. Basic Books Publishers: New York, NY. 1965
- [26] Chen A.G., Benrimoh D., Parr T., Friston K. A Bayesian account of generalist and specialist formation under the active inference framework. *Frontiers in Artificial Intelligence*. <https://doi.org/10.3389/frai.2020.00069>; 2020
- [27] Yufik Y. M. Understanding, consciousness and thermodynamics of cognition. *Chaos Solitons Fractals*. 2013; 55, 44-59.
- [28] Yufik Y.M. Gnostron: A framework for human-like machine understanding. *Proceedings. IEEE Symp. Computational Intelligence (SSCI)*, Bangalore, India 2019; 18-29.
- [29] Yufik Y. M. The understanding capacity and information dynamics in the human brain. *Entropy*, 2019; 21, 308, p. 1-38.
- [30] Yufik Y. M., Friston, K. J. Life and understanding: Origins of the understanding capacity in the self-organizing nervous system. *Frontiers Systems Neuroscience*. 2017; 10, 98.
- [31] Cattell, R. B. Intelligence: Its Structure, Growth and Action. Elsevier, NY. 1978.
- [32] Cattell R. B. 1971. Abilities: Their Structure, Growth, and Action. Houghton Mifflin, N.Y. 1971
- [33] Cattell R. B. Theory of fluid and crystallized intelligence: A critical experiment. *Journal Educational Psychology*. 1963; 54, 1, 1-22.
- [34] Harari, Y. N. Sapiens: A brief history of human kind. Harper Perennial. 2018.
- [35] Chomsky, N. Biolinguistic explorations: design, development, evolution. *International Journal Philosophical Studies*. 2007: 15, 1-21.
- [36] Hauser M.D., Chomsky N., Fitch, W.T. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science*, 2002; 298, 1569-1579.
- [37] Berwick R. C., Chomsky, N. Why only us: language and evolution. The MIT Press. 2017.
- [38] Caramelli 2008 Caramelli, D., Milani, L., Vai, S., Modi, A., Pecchioli, E., Girardi, M., Pilli, E., Lari, M., Lippi, B., Ronchitelli, A., Mallegni, F., Casoli, A., Bertorelle, B., Barbujani, G. 2008. A 28,000 years old Cro-Magnon mtDNA sequence differs from all potentially contaminating modern sequences. *PLoS ONE* 2008, 3(7): e2700.
- [39] Hennenberg, D. 1988 Decrease of human skull size in the Holocene. *Human Biology*, 1988, 60(3), p. 395-405
- [40] Calvin W.H. A brief history of the mind: From apes to intellect and beyond. Oxford University Press. NY. 2004
- [41] Smaers J. B., Soligo C. Brain reorganization, not relative brain size, primarily characterizes anthropoid brain evolution. *Proceedings Royal Society* 2013; B 280: 20130269
- [42] Deban S.M., Scales J.A., Bloom S.V., Easterling C.M., O'Donnell M.K., Jeffrey P., Olberding J.P. Evolution of a high-performance and functionally robust musculoskeletal system in

salamanders. *Proceedings National Academy Sciences*, 2020; 117 (19) p. 10445-10454

[43] Piaget J. *The Grasp of Consciousness: Action and Concept in the Young Child*. Taylor & Francis, NY. 1974

[44] Calvin W.H. *The Cerebral Code*. A Bradford Book. The MIT Press. 1996.

[45] Chomsky N. *Biolinguistic explorations: design, development, evolution*. *International Journal Philosophical Studies*. 2007; 15, p. 1-21.

[46] Piaget, J. *Success and Understanding*. Harvard University Press, Cambridge, MA. 1978

[47] Parr T., Friston, K. *The anatomy of inference: Generative models and brain structures*. *Frontiers Computational Neuroscience*. 2018; 12, 90

[48] Lear, J. *Aristotle The desire to understand*. Cambridge University Press. 1988.

[49] Rosen R. *Life Itself*. Columbia University Press. 1991.

[50] Haken, H. *Synergetics: An Introduction. Nonequilibrium Phase Transitions and Self- Organization in Physics, Chemistry and Biology*. Springer. 1996.

[51] Georgopoulos A.P. Lurito J.T., Petrides M., Schwartz A.B. Massey, J.T. *Mental rotation of the neuronal population vector*. *Science*, 1989; 243, 234-236.

[52] Georgopoulos A.P. Massey J.T. *Cognitive spatial-motor processes 1. The making of movements at various angles from a stimulus direction*. *Experimental Brain Research*. 1987; 65, p. 361-370

[53] Georgopoulos A.P., Taira, N., Lukashin, M. A. *Cognitive*

neurophysiology of the motor cortex. *Science*. 1993, 260, 47-52.

[54] Papo D. *Brain temperature: what it means and what it can do for (cognitive) neuroscientists*, 2013: arXiv:1310.2906

[55] Koestler, A. *The Act of Creation*. Arkana, London, UK. 1964.

[56] James W. 1950. *The Principles of Psychology*. Dover Publications, NY, USA. 1950

[57] Fritz J. B., David S., Shamma S. *Attention and dynamic, task-related receptive field plasticity in adult auditory cortex*. In: Cohen Y.E., Popper A.N., Fay R.R. (ed) *Neural Correlates of Auditory Cognition*. Springer. 2013; p.251-290

[58] Fritz, J. B., Elhilali, M., & Shamma, S. A. *Adaptive receptive field changes during detection of complex spectral targets*. *Journal of Neurophysiology*. 2007, 98, p. 2337-2346

[59] Fritz, J. B., Elhilali, M., & Shamma, S. A. *Active listening: Task-dependent plasticity of receptive fields in primary auditory cortex*. *Hearing Research*, 2005; 206, 159-176.

[60] Fritz J. B., Shamma S. A., Elhilali M., & Klein, D. J. *Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex*. *Nature Neuroscience*. 2003; 6, p. 1216-1223.

[61] Atiani S., Elhilali M., David S. V., Fritz J. B., Shamma, S. A. (2009). *Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields*. *Neuron*. 2009; 61, p. 467-480.

[62] Duncan, J. *An adaptive coding model of neural function in prefrontal*

cortex. *Nature Reviews Neuroscience*, 2001; 2, 820-829.

[63] Stephen V. D., Fritz J.B., Shihab A. S. *Proceedings National Academy Sciences*, 2012; 109 (6) p. 2144-2149

[64] Christie S.T., Schrater P. Cognitive cost as dynamic allocation of energetic resources. *Frontiers in Neuroscience*, <https://doi.org/10.3389/fnins.2015.00289>, 2015

[65] Lieder F., Griffiths T.L. Resource--rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral Brain Sciences*, 2019; 43, e1:1-60.

[66] Shaw, M.L., Shaw P. Optimal allocation of cognitive resources to spatial locations. *Journal Experimental Psychology*. 1977; 3(2), p.201-211.

[67] Dehaene S., Kerszberg M., Jean-Pierre Changeux J-P. A neuronal model of a global workspace in effortful cognitive tasks *Proceedings National Academy Sciences* 1998; 95 (24), p. 14529-14534

[68] Kouneiher F., Charon S., Koechlin E. Motivation and cognitive control in the human prefrontal cortex. *Nature Neuroscience*, 2009; 12 (7), p. 939-947.

[69] Kurzban R., Duckworth A., Kable, J.W., Myers, J. An opportunity cost model of subjective effort and task performance. *Behavioral Brain Sciences*, 2013. 36, p. 661-726

[70] Duncan J., Seltz R. J., Kolodny J., Bor D., Herzog H., Ahmed A., Newell F. N., Emslei H. A neural basis for general intelligence. *Science*, 2000; 289, 5478, p. 457-460

[71] Anderson M. L. Neural reuse: A fundamental organizational principle of the brain. *Behavioral Brain Sciences*, 2011; 33, p. 245-313.

[72] Bassett D. S., Greenfield D.L., Meyer-Lindenberg A., Weinberger, D. R., Moore, S. W., Bullmore, E.T. Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Computational Biology*, 2010; 6, 4, e1000748

[73] Cherniak C., Mokhtarzada Z., Rodrigues-Esteban R., Changizi, K. Global optimization of cerebral cortex layout. *Proceedings National Academy Sciences*. 2004; 101, p. 1081-1086.

[74] Attwell D., Laughlin S. B. An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab*. 2001; 21, p. 1133-1145.

[75] Raichle M. E. , Mintun M. A. Brain work and brain imaging. *Annual Reviews Neuroscience*. 2006; 29:449-476.

[76] Rolfe et al, 1997; Rolfe, D. F., Brown GC. Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiology Reviews*, 1997; 77(3), 731-758.

[77] Shulman R. G., Rothman D. L., Behar K. L., Hyder, F. Energetic basis of brain activity: implications for neuroimaging. *Trends Neurosci*. 2004; 27, 489-495.

[78] Sengupta B., Stemmler M. B., Friston, K. J. Information and efficiency in the nervous system—A synthesis. *PLoS Computational Biology*, 2013; 9, 7: e1003157.

[79] Zhu X.-H., Qiao H., Du F., Xiong Q., Xiao L., Zhang X., Ugurbil K., Chen W. Quantitative imaging of energy expenditure in human brain, *Neuroimage*, 2012, 60 (4), p. 2107-2117.

[80] Tomasi, D., Wang, G.-J., Volkow, N. D. 2013. Energetic cost of brain functional connectivity. *Proceedings National Academy Sciences*, 2013; 110 (33), p. 13642-13647.

- [81] Rae C., Scott R. B., Lee M., Simpson J. M., Hines N., Paul C., Anderson M., Karmiloff-Smith A., Styles, P., Radda, G.K. 2003. Brain bioenergetics and cognitive ability. *Developmental Neuroscience*, 2003, 25 (5) p. 324-331.
- [82] Pepperell R. Consciousness as a physical process caused by the organization of energy in the brain. *Frontiers in Psychology*, 2019. <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02091/full>
- [83] Markman A.B., Otto, A.R. 2011. Cognitive systems optimize energy rather than information. *Behavioral Brain Science*, 2011; 34 (4), 207.
- [84] Niven J. E., Laughlin S. B. Energy limitation as a selective pressure on the evolution of sensory systems. *Journal Experimental Biology*, 2008; 211, p. 1792-1804.
- [85] Laughlin S.B., Attwell D. Neural energy consumption and representation of mental events. In: Shulman R.G., Rothman D. L.(eds.) *Brain energetics and neuronal activity: Applications to fMRI and medicine*. Wiley, 2004; p.11-124.
- [86] Perin R., Berger T. K., Markram H. A synaptic organizing principle for cortical neuronal groups. *PNAS* 2011. 108 (13), p. 5419-5424
- [87] Huang H.J., Kram R., Ahmed, A.A. Reduction of metabolic cost during motor learning of arm reaching dynamics. *Journal of Neuroscience*, 2012; 32960, p. 2182-2190.
- [88] Sparrow W. A., Lay B. S., O'Dwyer N. J. Metabolic and attentional energy costs of interlimb coordination. *Journal of Motor Behavior*, 2007; 39 (4), p. 259-275
- [89] Sparrow W. A., Newell K. 1998. Metabolic energy expenditure and the regulation of movement economy. *Psychonomic Bulletin & Review*. 1998; 5 (2), p. 173-196
- [90] Sparrow W. A., Newell K. M. Energy expenditure and motor performance relationships in humans learning a motor task. *Psychophysiology*, 1994; 31, 338-346.
- [91] Sparrow W. A., Irizarry-Lopez V. M. 1987. Mechanical efficiency and metabolic cost as measures of learning a novel gross-motor task. *Journal Motor Behavior*. 1987; 19, 240-264.
- [92] Prilutsky, B. I. , Zatsiorsky, V. M. 2002. Optimization-based models of muscle coordination. *Exercise Sport Science Reviews*. 2002; 30, 32-38s
- [93] O'Dwyer N. J., Neilson P. Metabolic energy expenditure and accuracy in movement: Relation to levels of muscle and cardiorespiratory activation and the sense of effort. In: Sparrow W. A. (ed.), *Energetics of Human Activity*. Human Kinetics, Champaign, 2002; p. 1-42
- [94] Westbrook A., Braver T. S. 2015. Cognitive effort: A neuroeconomy approach. *Cognitive, Affective Behavioral Neuroscience*. 2015; 15, p. 395-415.
- [95] Hebb, D.O. *The organization of behavior*. Wiley & Sons, N.Y. 1949.
- [96] Edelman G. M., Gally J.A. 2013. Reentry: a key mechanism for integration of brain function. *Frontiers in Integrative Neuroscience*. 2013; 7, 63
- [97] Edelman G. *Neural Darwinism. The theory of neuronal group selection*. Basic Books, NY. 1987.
- [98] Edelman, G. *Bright air, brilliant fire: On the matter of the mind*. Basic Books, N.Y. 1992.
- [99] Thoroughman K., Shadmehr R. Learning of action through adaptive

- p>combination of motor primitives. Nature, 2000. 407, p. 742-747.
- [100] Vyshedskiy, A. 2019. Neuroscience of imagination and implications for human evolution. Current Neurobiology 2019; 10 (2) p. 89-109.
- [101] Lin L., Osan R. Tsien J.Z. Organizing principles of real-time memory encoding: Neural clique assemblies and universal neural codes. Trends Neuroscience. 2006; 29, 48-57
- [102] Gelfand I. M, Latash M. L. On the problem of adequate language in movement science. Motor Control. 1998; 2, 306-313
- [103] Gelfand, I. M., Latash M. L. On the problem of adequate language in biology. In: Latash M. L. (ed). Progress in Motor Control. vol. 2: Structure-Function Relations in Voluntary Movement. Human Kinetics, Urbana, IL, 2002. p. 209-228.
- [104] Latash M. L. Biomechanics as a window into the neural control of movement. Journal Human Kinetics 2016; 52, 7-20.
- [105] Latash M. L. Synergy. Oxford University Press, N. Y. 2008.
- [106] Latash M. L. , Scholz J. P. , Schöner G. Toward a new theory of motor synergies. Motor Control. 2007; 11, 276-308
- [107] Georgopoulos, A.P. Lurito, J.T., Petrides, M., Schwartz, A.B. Massey, J.T. 1989. Mental rotation of the neuronal population vector. Science, 1989: 243, p. 234-236.
- [108] Georgopoulos A.P., Massey J.T. Cognitive spatial-motor processes 1. The making of movements at various angles from a stimulus direction. Experimental Brain Research. 1987: 65, p. 361-370.
- [109] Georgopoulos A.P., Taira M., Lukashin, A. Cognitive neurophysiology of the motor cortex. Science 1993: 260, p. 47-52.
- [110] Georgopoulos A.P., Kettner, R.E., Schwartz, A.B. Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. Journal Neuroscience 8, 2928-2937.
- [111] Latash M.L. Neurophysiological Basis of Movement. Human Kinetics. 2008.
- [112] Vaz A.P., Wittig Jr. J.H., Inati S.K., Zaghoul K.A. Replay of cortical spiking sequences during human memory retrieval. Science, 2020: 367 (6482), p. 1131-1134
- [113] Graybiel A.M. The basal ganglia and chunking of action repertoires. Neurobiology of Learning and Memory. 1998: 70, p. 119-136
- [114] Jin X., Tecuapetla F., Costa R. M. 2014. Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences, Nature Neuroscience. 2014. 17 (3), p. 423-430.
- [115] Schmitt L., Wimmer R., Nakajima M., Happ M., Mofakham S., Halassa M.M. Thalamic amplification of cortical connectivity sustains attentional control. Nature 2017: 545, p. 219-223
- [116] Miller R. Cortico-thalamic interplay and the security of operation of neural assemblies and temporal chains in the cerebral cortex. Biological Cybernetics 1996: 75, 263-275.
- [117] Braitenberg V., Heck D., Sultan F. et al. The detection and generation of sequences as a key cerebellar function: Experiments and theory. Behavioral Brain Sciences 1997: 20 (2) p. 229-277.
- [118] Bressler S. L., Kelso J.A. Coordination dynamics in cognitive

neuroscience. *Frontiers in Neuroscience* 2016, <https://doi.org/10.3389/fnins.2016.00397>

[119] Bressler S. L., Kelso J.A. Cortical coordination dynamics and cognition. *Trends in Cognitive Neuroscience* 2001;5(1) p. 26-36.

[120] Bressler S. L., Tognoli, E. Operational principles of neurocognitive networks. *International Journal Psychophysiology*. 2006: 60, 139-148.

[121] Kelso J.A.S., Dumas G., Tognoli E. Outline of a general theory of behavior and brain coordination. *Neural Networks*. 2013: 37, 120-131.,

[122] Edelman G. The remembered present: A biological theory of consciousness. Basic Books, NY. 1989.

[123] Edelman G. Bright air, brilliant fire: On the matter of the mind. Basic Books, N.Y. 1992.

[124] Edelman G., Changeaux, J-P. The Brain. Transaction Publishers. NY. 2001.

[125] Braitenberg, V. Cell assemblies in the cerebral cortex. In: Heim R., Palm G. (eds) *Theoretical approaches to complex systems. Lecture Notes in Biomathematics*, 21. Springer, Berlin, 1978: p. 171-188.

[126] Feldman, A.G., Goussev, V., Sangole, A., Levin, M.F. Threshold position control and the principle of minimal interaction in motor actions. *Progress in Brain Research* 2007: 165, p. 267-281.

[127] Feldman, A. G., Ostry, D.J., Levin, M. F., Gribble, P. L., Mitnitski, A. B. Recent tests of the equilibrium-point hypothesis (λ model). *Motor Control* 1998: 2, p. 189-205

[128] Feldman A.G., Levin M.F. The equilibrium-point hypothesis – past,

present and future. In: Sternad D. (eds) *Progress in Motor Control. Advances in Experimental Medicine and Biology*, Springer, Boston, 2009. p. 699-726

[129] Doyle J., Ford D. Mental models concepts for system dynamics research. *System Dynamics Review*. 1998: 14 (1) p. 3-29

[130] Krivov S.V., Karplus M. Hidden complexity of free energy surfaces for peptide (protein) folding. *Proceedings National Academy Sciences*. 2004: 101 (41) p. 14766-14770

[131] Bryngelson J. D., Onuchic J. –N., Socci N. D., Wolynes P. G. Funnels, pathways and the energy landscape of protein folding: A Synthesis. *Proteins* 1995: 2 (3) p. 167-195

[132] Abeles M., Bergman H., Gat I., Meilijson I., Seidemann E., Tishby N., Vaadi E. Cortical activity flips among quasi-stationary states. *Proceedings National Academy of Science*. 1995: 92, p 8616-8620

[133] Miller E. K., Cohen J. D. 2001. An integrative theory of prefrontal function. *Annual Reviews Neuroscience*. 2001: 24, 167-202

[134] Tang E., Giusti G., Baum G.L., Gu S., , Pollock E., Kahn A.E., Roalf D.R. Moore T.M., Ruparel K., Gur R.C., Gur R.E., Satterthwaite T. D., Bassett D.S. Developmental increases in white matter network controllability support a growing diversity of brain dynamics. *Nature Communications* 2017: 18, 1252

[135] Ezaki T., Watanabe T., Ohzeki M., Masuda N. Energy landscape analysis of neuroimaging data. *Philosophical Transactions Royal Society*. 2017: A.37520160287

[136] Watanabe T., Hirose S., Wada H., Imal Y., Machida T., Shirouzu I., Konichi S., Miyashita Y., Masuda N., Energy landscapes of resting state brain

networks. *Frontiers Neuroinformatics*, 2014 <https://doi.org/10.3389/fninf.2014.00012>

[137] Gu, S., Cieslak, M., Baird, B., Muldoon, S.F., Grafton, S.T., Pasqualetti, F., Bassett, D.S. The energy landscape of neurophysiological activity implicit in brain network structure. *Scientific Reports*. 2018: 8, 2507, 1-15.

[138] Deco G., Jirsa V.K., McIntosh V.R. Resting brains never rest: Computational insights into potential cognitive architectures. *Trends Neurosciences*. 2013: 36 (5) p. 268-274

[139] Wei Y., de Lange S. C., Scholtens L. H., Watanabe K., Ardesch D. J., Jansen P. R., Savage J. E., Li L., Preuss T.M., James K., Rilling, J. K., Posthuma, D., van den Heuvel M.P. Genetic mapping and evolutionary analysis of human-expanded cognitive networks. *Nature Communications*, 2019: 10, 4839

[140] Raichle M. E. , Mintun M. A. Brain work and brain imaging. *Annual Reviews Neuroscience*. 2006: 29, p. 449-476.

[141] Seeley, W. W. The Salience Network: A neural system for perceiving and responding to homeostatic demands. *Journal of Neuroscience*, 2019: 39 (50) p. 9878-9882

[142] Nelson, S.M., Dosenbach, N.U.F., Cohen, A. L., Wheeler, M.E., Schlaggar, B.L., Petersen, S.E. Role of the anterior insula in task-level control and focal attention. *Brain Structure Functions*, 2010: 214, 5-6, p. 669-680.

[143] Smith, S.M., Fox, P.T., Miller, K. L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Angela R. Laird, A.R., Beckmann, C.F. Correspondence of the brain's functional architecture during activation and rest. *PNAS* 2009: 106, p. 31-42

[144] Klee, M.R., Pierau, F.-K., Faber, D.S. Temperature effects on resting potential and spike parameters of cat motoneurons. *Experimental Brain Research* 1974: 19, p. 478-492.

[145] Beggs, J. M., Plentz, D. Neuronal avalanches in neocortical circuits. *Journal Neuroscience*, 2003: 23 (35) p. 1167-11177.

[146] Goulden, N., Khusnulina, A., Davis, N.J., Bracewell R.M., , Bokde, A.L., McNulty, J.P., Mullins, P.G. The salience network is responsible for switching between the default mode network and the central executive network: Replication from DCM. *Neuroimage*, 2014: 99, p. 180-190.

[147] Amunts K., Schleicher, A., Zilles, K. Cytoarchitecture of the cerebral cortex. *NeuroImage* 2007: 37, p. 1061-1065

[148] Zilles, K., Amunts, K. 2015. Anatomical basis for functional specialization. In: Uludag, K., Ugurbil, K. & Berliner, L. (eds.). *From nuclear spins to brain functions*. Springer, 2015 p. 27-66.

[149] Hipolito I., Ramstead M., Convertino L., Bhat A., Friston K., Parr T. Markov Blankets in the Brain. 2020arXiv:2006.02741 [**q-bio.NC**]

[150] Buckner, R.L., Krienen, F.M. The evolution of distributed association networks in the human brain. *Trends in Cognitive Sciences*. 2013:17 (12), p.648-665.

[151] Laurent, G. What does 'understanding' mean?. *Nature Neuroscience*, 2000: 3, 1211

[152] Hough A.R., Gluck K.A. The understanding problem in cognitive science. *Advances in Cognitive Systems*, 2019: 8, 13-32

- [153] Waltz J.A., Knowlton B.J., Holoyak K.J., Boone K.B., Mishkin F.S., Santos M., Thomas C.R., Miller B.L. A system for relational reasoning in human prefrontal cortex. *Psychological Science*, 1999: 10 (2), p. 119-125.
- [154] Krawczyk, D.C. The cognition and neuroscience of relational reasoning. *Brain Research*, 2012:1428, 13-23.
- [155] Clement J. Creative model construction in scientists and students. Springer, The Netherlands 2008.
- [156] Clement, J. J. The role of imagistic simulation in scientific thought experiments. *Topics in Cognitive Science*. 2009:1, p. 686-710
- [157] Hegarty, M. Mechanical reasoning by mental simulation. *Trends Cognitive Science*. 2004: 8(6), p. 280-285.
- [158] Kanai, R., Chang A., Yu, Y., de Abril, I. M., Biehl, M., Guttenberg, N. 2019. Information generation as a functional basis of consciousness. *Neuroscience of Consciousness*, 2019: 5, 1 niz016.
- [159] Treisman, A. Focused attention in the perception and retrieval of multidimensional stimuli. *Perception & Psychophysics*, 1977: 22 (1) p. 1-11.
- [160] Crick, F. Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings National Academy Sciences*. 1984: 81, 4586-4590.
- [161] Rizzolatti, G., Arbib, M. 1998. Language within our grasp. *Trends Neuroscience*. 1998: 21 (5) p. 188-194.
- [162] Friston, K. J., Parr, T., Yufik, Y., Sajid, N., Catherine J. Price, C. J., Holmes, E. Generative models, linguistic communication and active inference. *Neuroscience and Biobehavioral Reviews*. 2020: 118, p. 42-64.
- [163] Pulvermuller, F. Words in the brain's language. *Behavior and Brain Science*. 1999: 22, p. 253-336.
- [164] Pulvermuller, F. Hebb's concept of cell assemblies and the psychophysiology of word processing. *Psychophysiology* 1996: 33, p. 317-333
- [165] Barsalou, L. W. Grounded cognition. *Annual Reviews Psychology*, 2008: 59, p. 617-645
- [166] Boulenger, V., Hauk, O., Pulvermuller, F. Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cerebral Cortex* 2009: 19, p. 1905-1914
- [167] Nelson, M. J., Karouib, I. E., Giberc, K., Yang, X., Cohen, L., Koopmanf, H., Cashc, S.C., Naccacheb, L., Haleg, J.T., Palliera, C., Dehaene, S. Neurophysiological dynamics of phrase-structure building during sentence processing *Proceedings National Academy of Sciences*, 2017: E3669–E3678
- [168] Horwitz, B.; Rumsey, J.M.; Grady, C.L. The cerebral metabolic landscape in autism; intercorrelations of regional glucose utilization. *Arch. Neurol*. 1988, 45, 749-755.
- [169] Just, M.A.; Cherkassky, V.L.; Keller, T.A.; Minshew, N.J. Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity. *Brain* 2004, 127,1811-1821.
- [170] McAlonnan, G.M.; Cheung, V.; Cheung, C.; Suckling, J.; Lam, J.Y.; Tai, S.; Yip, L.; Murphy, D.G.M.; Chua, S.E. Mapping the brain in autism: A voxel-based MRI study of volumetric differences in autism. *Brain* 2005, 128, 268-276.
- [171] Parsey, C. M., Schmitter-Edgecombe, M. Quantitative and

qualitative analyses of the clock drawing test in mild cognitive impairment and Alzheimer disease: Evaluation of a modified scoring system. *Journal of Geriatric Psychiatry and Neurology*. 2011; 24(2), p. 108-118.

[172] Smith, E., Morowitz, H. J. *The Origin and Nature of Life on Earth*. Cambridge University Press, U.K. 2016.

[173] Toyabe, S., Sagawa, T., Ueda, M. Muneyuki, E., Sano, M., 2010. Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nature Physics* 2010: 6, p. 988-992

[174] Penrose, R. On understanding understanding. *International Studies in the Philosophy of Science*, 1997: 11 (1) p. 7-20