UNDERSTANDING CAPACITY IN THE HUMAN AND THE MACHINE

Yan M. Yufik

Abstract—**For millennia, questions concerning human understanding capacity have been relegated to philosophy, but became an issue of technical urgency recently, due to rapid advancements in machine intelligence. This paper summarizes a theory of neuronal mechanisms underlying understanding followed by outlining a framework for simulating such mechanisms in computational artifacts.**

Keywords—**understanding,mental modeling, neuronal packets**

I.   Introduction

This introductory section defines terms and summarizes the key ideas in the theory of understanding [1], [2]. The theory associates intelligence with adaptive regulatory mechanisms serving to organize interaction between organism and environment in a manner conducive to survival. Understanding is an advanced form of such regulatory mechanisms that is unique to the humans. Adaptive regulation is centered on the learning capacity which is ubiquitous to all species and boils down to optimizing organism-environment interaction based on the prior interaction history.   The mechanism works (optimizes responses) when the current conditions are similar to some of those encountered in the past, and fails when conditions are novel and unfamiliar. Understanding complements learning, by allowing construction of near-optimal responses in novel situations having no precedents in the past history.   The capability is due to mental modeling which, in a sense, compensates for the lack of past experiences without the need for recreating them. During learning, experiences (interactions with objects in the environment) are registered in neuronal structures which get re-activated when encountering similar objects and conditions. Understanding involves self-directed (decoupled from sensory inputs) activation and manipulation of neuronal structures. Stated differently, mental modeling allows manipulation of mental objects decoupled from the physical objects they represent. A more precise definition follows.

Physical objects. Physical objects have three definitive attributes: they are a) distinct, i.e. occupy volumes in space

Yan M. Yufik
Virtual Structures Research, Inc.
USA

delineated by distinct boundaries, b) preserve their self-identity, i.e. change their properties and admit a certain amount of changes without losing their self-identity and c) can interact with other objects, i.e., form relations with other objects imposing some form of coordination on the object's behavior (changes over time).

b) Mental objects. Mental objects are of two types: neuronal structures and virtual objects. Neuronal structures reside in the neuronal space and are counterparts (representations) of physical objects residing in the physical space. Neuronal structures give rise to virtual objects which are organized sensory groupings, or images (visual, acoustic, tactile, etc.) Virtual objects reside in the mental space.

c) **Mental modeling.** Mental modeling involves operations on neuronal structures underlying the experience of "thinking," that is, manipulating virtual objects in the mental space.

The theory of understanding is centered on the notion of "**neuronal packets**" which are Hebbian assemblies surrounded by boundary energy barriers. Neuronal packets are counterparts of physical objects: boundary energy barriers make them distinct and thus capable of representing distinct objects, their behavior and relations. The next section defines these notions in greater detail.

I.   Neuronal packets and mental modeling.

Energy barriers make neuronal packets quasi stable, i.e., capable of withstanding entropic erosion and changes in conditions both inside and outside packet boundaries. Crucially, they also make packets discernible, i.e., separate and distinct from each other, and amenable to being combined into functional groupings (mental models) that can have properties different from those of the constituent packets. Stability, perseverance and discernibility of neuronal packets underlie the experience of the external world being populated

with bounded entities that resist disintegration and maintaining their self-identity under changing conditions.

Packets respond to variations in the stimulation flow, by varying patterns of excitation-inhibition in their constituent neurons. Varying patterns in packet **a** underlie the experience of behavior variation in the corresponding physical object A (i.e., changes in A admissible under its self- identity). If population vectors are associated with packets (Packet Response Vectors, or PRVs) behavior of objects can be represented as PRV movements, as shown in Figure 1.
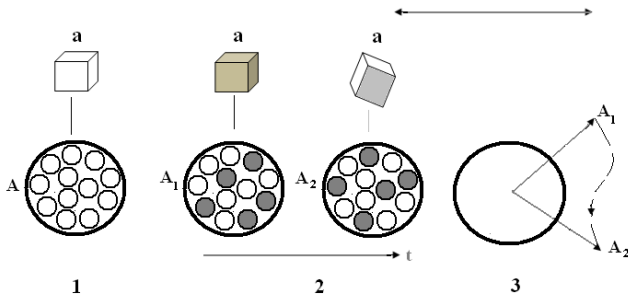

Figure 1. 1) Neuronal representations of external objects are constituted by neuronal packets. 2) Changes in the external objects (i.e., behavior) cause variations of excitation-inhibition patterns in the corresponding packets. 3) Behavior can be expressed as trajectories of Packet Response Vectors (PRVs).

Figure 1 conveys one of the key hypotheses in the theory of understanding: behavior of physical objects is represented as rotation of packet vector in a packet or group of packets underlying perception of those objects. Figure 1.1 depicts neuronal packet as a complex, or manifold comprising multiple neurons while Figure 1. 3 depicts the same packet as a holistic unit. Forms of object perception in Figure 1.1 and Figure 1.3 are mutually exclusive, not unlike in handling a book: one can either open a book and examine its contents or close the book and temporarily loose access to the contents. Another example: one can experience apple as a rich sensory
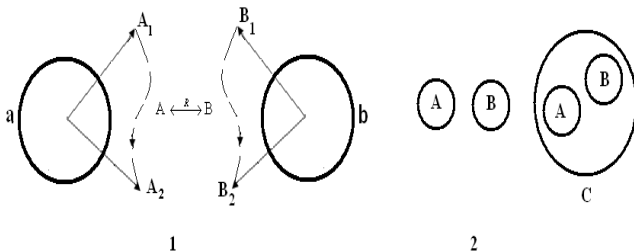

Figure 2. 1) Juxtaposing packets and detecting some form of inter-packet coordination establishes *relations* (particular forms of behavior dependency) between the corresponding external objects. 2) Encountering objects A and B can be separated by large time periods and multiple other encounters. Composition brings together packets **a** and **b** in a composite structure capturing relation between A and B and thus combining them into a relational unit C that is amenable to inclusion into other compositions. In this way, iterative cycles of composition produce nested models capturing dependencies between expanding varieties of objects and object groupings.

grouping (color, taste, etc.) or as a unit defined by the way it is grouped with and behaves respective other units, as in "putting apple on the table". Reversible unit-manifold transitions (dubbed 'packet enfolding/packet unfolding') is the single most fundamental operation underlying understanding. Figure 2 depicts construction of mental models.

Detection of inter-packet coordination requires neurons sensitive to particular forms of dependency in the movement of packet vectors (i.e., correspondences in the behavior (excitation-inhibition patterns) in different packets). Accordingly, the theory postulates existence of regulatory neurons that respond to or can be formed into groups responding to different types of pattern correspondence. Eliciting responses from such neurons underlies the experience of forming models capturing relations between objects. Self-initiated activation of coordination neurons produces coordinated activation/inhibition patterns in the packets experienced as 'imagining' the relations.

II. Gnostron: a framework for machine understanding.

In the gnostron framework, cognitive process is defined as probabilistic optimization of neuronal resources [3], [4], [5], [6], [7], [8] undergoing two stages: formation of neuronal packets (learning), followed by constructing and manipulating packet compositions (modeling).

During learning, neurons are drawn from pool $X = x_1, x_2, ..., x_N$ and allocated to streams of stimuli drawn from a finite set $Z = z_1, z_2, ..., z_M$. Pool X has finite longevity allowing a succession of L stimulation episodes $Z_1, Z_2,...,Z_L$ comprising varying stimuli combinations of $Z_q = \{z_p, z_b,..., z_w\}$. Combinations in different episodes can overlap, $Z_q \cap Z_v = \varnothing$, $q \neq v$. Neurons in X are defined by their response vectors, as shown in Figure 3.
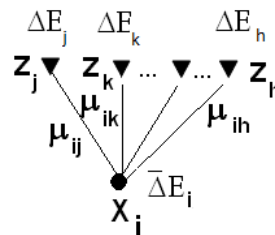
Figure 3 . Response range of neuron x_i includes stimuli z_j, z_k,…, z_h. Firing x_i in the presence of z_i is rewarded with ΔE_i amount of energy with probability μ_{ij}, and incurs energy cost in the amount $\overline{\Delta E}_i$. Response vector of x_i is defined in M-dimensional stimuli space and comprises probability values $\vec{x}_i = (\mu_{ik}, \mu_{ij}, ..., \mu_{ir})$.

Neurons respond to different stimuli (by probabilistic firing) and, reciprocally, stimuli respond to different neurons (by probabilistic rewards), as shown Figure 4.
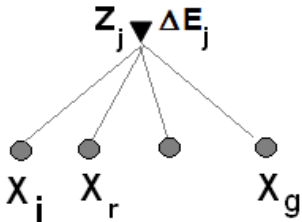

Figure 4 Stimulus z_j responds to collective firing of neuronal group, by emitting energy reward ΔE_j.
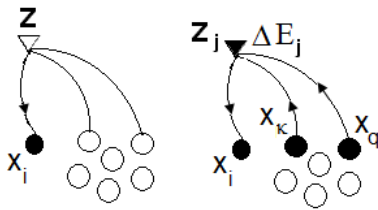
Figure 5 illustrates generic problem facing gnostron

Figure 5. Blob Z caused firing of neuron x_i: What other neurons need to be fired in order to extract energy reward from Z? Successful firing obtains reward and identifies the blob as stimulus z_j.

Neuron-stimuli allocations are carried out within windows of fixed duration and capacity, as shown in Figure 6.
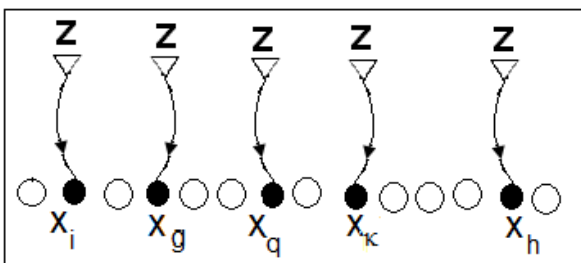

Figure 6. Neurons are allocated to stimuli to maximize the cumulative reward. Allocations of initially uncommitted neurons (unshaded circles) are based on the configuration of responses caused by the yet unidentified stimuli (blobs). Reward maximization is concomitant with maximizing uncertainty reduction (reliable identification of the stimuli).

Pool's survival (for the duration of its longevity limit) requires allocation of neuronal resources in every episode yielding net energy supplies (the sum of intakes minus internal expenditures) equal or exceeding some threshold value E_t, as expressed in Equation 1 below.

$$F_l(\delta_{ij}) = \sum_{i=1}^{m} (\Delta E_i (1 - \prod_{j=1}^{n} \varepsilon_{ji}(\delta_{ji})) - \sum_{j=1}^{n} \overline{\Delta E}_j(\delta_{ji}) \geq E_t$$

here $\delta_{ij} = 1$ if neuron x_i is allocated (tuned and fired) to z_j, and $\delta_{ij} = 0$ if otherwise; $\varepsilon_{ij} = 1 - \mu_{ij}$; m is the number of stimuli and n is the number of neurons participating in the episode.

Energy intake declines as a result of misfiring, i.e., expenditures are not compensated. The first term in the equation defines cumulative rewards yielded by the allocation while the second term defines cumulative expenditures (e.g., cost of firing). Evolution towards increased thermodynamic efficiency requires minimization of internal expenditures and maximization of energy inflows during the pool's lifetime, as indicated in Equation 2.

$$\sum_{L} F_l(\delta_{ij}) \rightarrow \max \qquad (2)$$

Survival requirements (Equations 1 and 2) are satisfied by forming neuronal groups in the pool $X^S = \{x_j, x_h, ...,x_t\}$ such that, whenever any neuron in the group happens to respond, all the other group members become candidates for allocation. More precisely, response vectors are associated with the groups allowing treating them as units allocated to stimuli in exactly the same fashion as the individual neurons

$$\vec{X}^s = (\mu_j^s, \mu_h^s, ..., \mu_t^s)$$

$$\mu_h^S = 1 - \prod_{h \in n_h^S} \varepsilon_h^S , \quad n_h^S \in n^S \qquad (3)$$

here n^S is the number of neurons in X^S.

Reversible transition between groups (treated as units) and their constituents is illustrated in Figure 7.
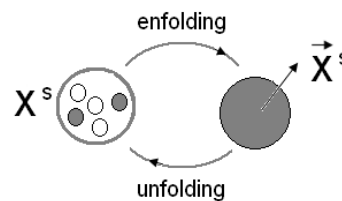

Figure 7. For allocation purposes, neuronal group can be treated, alternatively, as a unit (packet) or as a composition of lower-level units (neurons).

The allocation process alternates between allocating groups and allocating their constituents, as shown in Figure 8.
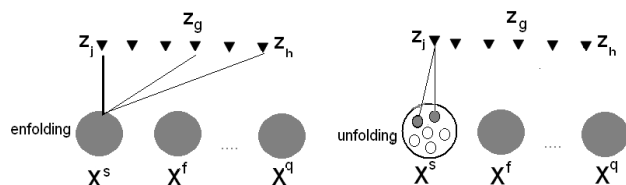
Figure 8. 1) Neuronal groups are allocated as units (group X^S is allocated to z_j). 2) Subsequent allocation of individual neurons in X^S is constrained by the preceding group allocation (neurons in X^S responsive to z_j are all tuned to z_j).

As configurations of the stimuli vary, group allocations change accordingly, as shown in Figure 9.
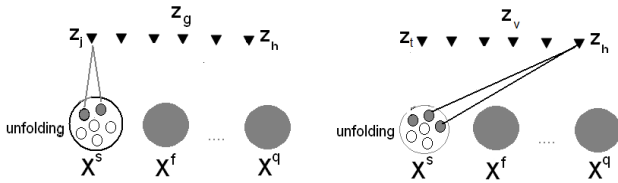

Figure 9. Group X^S is allocated differently as the stimuli configuration changes from 1 to 2. Involvement of different neurons in X^S corresponds to rotation of group vector $\vec{X}^S$.

Finally, establishing inter-packet coordination (constructing mental models) imposing order in selecting groups and constrains selection of resources (neurons) within the groups, as shown in Figure 10.
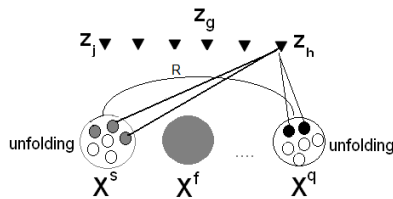

Figure 10. Establishing relation R between groups X^S and X^q causes resource allocation process to alternate between the groups while constraining selection of neurons within the groups.

Operating gnostron involves defining coordination patterns ("A pushes B", "A rests on top of B," etc.) and defining constraints for applying them within the domain of interest.

It has been shown that grouping by packets (i.e., via forming associative networks partitioned into internally cohesive and externally weakly coupled neuronal groups) is near-optimal, i.e., maximizes rewards while minimizing the number of unproductive allocations. Packet mechanism was shown to yield near-optimal results and orders of magnitude reduction in the amount of computation in large-scale resource optimization tasks (e.g., target recognition) [9], [10].

REFERENCES

[1] Y.M. Yufik, Understanding, consciousness and thermodynamics of cognition. Chaos, Solitons & Fractals 2013, 55, pp. 44–59

[2] Y.M. Yufik, Y. M. 1998. Virtual Associative Networks: A Framework for Cognitive Modeling", in: K. Pribram (ed.), Brain and Values, LEA, New Jersey, pp. 109-177,

[3] E. Lawler, Combinatorial optimization: Networks and matroids. Dover Publications, N.Y. 1976.

[4] C.H. Papadimitriou, K. Steiglitz, Combinatorial optimization: Algorithms and complexity, Prentice Hall, New Jersey, 1982

[5] E. Suhir, Applied probability for engineers and scientists, McGraw-Hill, 1997.

[6] I. Ushakov, Optimal resource allocation, Wiley, 2013.

[7] T. Ibaraki, N. Katoh, Resource allocation problems: Algorithmic approaches. The MIT Press, 1988.

[8] Y. M. Yufik, Probabilistic resource allocation systems with self-adaptive capabilities, U.S. patent 5,794, 224, 1998.

[9] Y.M. Yufik, R. Malhotra, Information blending in virtual associative networks: A new paradigm for sensor integration. Int. J. Art. Intel. Tools. 1999, 8, 3, pp. 275-290.

[10] Y. M. Yufik, T. Sheridan, Swiss Army knife and Ockham's razor: Modeling and facilitating operator's comprehension in complex dynamic tasks, IEEE Transactions on SMC, Part A: Systems and Humans, 2002, 32, 2, pp. 185-199,