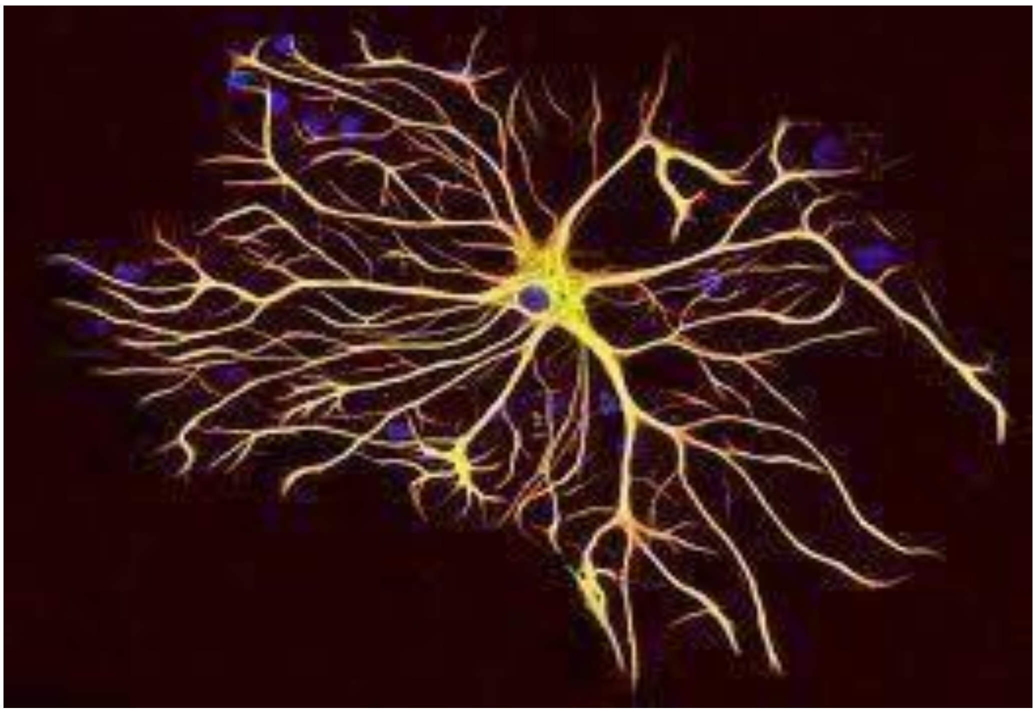
Speaking Brains <>

Collective Papers on AI, Neuro-circuitry and Basal Ganglia Grammar *Draft 2023*

Joseph Galasso

OSF | Joseph Galasso



Speaking Brains <> Collective Papers on AI, Neuro-circuitry and Basal Ganglia Grammar

Joseph Galasso
California State University, Northridge

The most compelling evidence to date for involvement of the Basal Ganglia (BG) (Basal Ganglia Grammar) in natural language comes to us from theoretical *movement operations* (nested dependency, distant binding and trace-theory). This implication of BG overlaps with well-established evidence showing Broca's involvement with movement. Dual pathways are a marked characteristic of BG insofar that in cascading down-stream neural networks, both direct as well as indirect paths affect admixed neuronal populations from multiple cortical areas. A tentative proposal may suggest that any notion of duality at the subcortical level may have the ability to simulate what we know of local vs distant binding dependencies as found in Dual Mechanism Model accounts of natural language. A theoretical (meta)-synthesis which seeks to connect what we know of Natural Language (NL) with current trends in AI/Transformers may offer us a potential merging of what has up until now been two quite disparate underlying systems. If we assume that NL systems mirror what we find in Parallel Distributed Processing (PDP) across neural networks—and via extension be applicable to any putative AI/Transformer-to-NL corollary—then, by definition, some component of the PDP would necessarily entail a *capacity-state* which corresponds to *concepts, symbols and categorial rules*—i.e., *real recursive-based* prerequisites for natural language which up until now have been sidelined in the implementation of AI modeling.

ISBN 9783969392065. **LINCOM Studies in Neurolinguistics 04**. 96pp. 2024.

Preface

In a 1995 book entitled *Speaking Minds,* editors Peter Baumgartner and Sabine Payr put together a fascinating series of interviews with twenty of the most eminent cognitive scientists of the twentieth century. Out of these interviews emerged just how deep-seated and explicitly contentious animosities ran between some of these great minds, and showcased just how it was inevitable that two camps on AI would eventually splinter. Like two emerging phoenixes out of the cognitive ashes of unfulfilled promises, this Janus-headed monster would take on almost religiously zealous overtones and contempt for one another, as both sides would attempt to explain away the other's respective shortcomings in what was at the time thought of as an emerging field which held so much promise for future AI. While the interviews in 1995 seemed new and nuanced by today's standards, the debates themselves have much earlier antecedents dating back to pioneers such as Donald Hebb ('neurons which fire together wire together'), and pre-war brilliant polymaths such as von Neumann and Turing himself (the Turing Test)—all of whom fostered the famous post-war debates between Marvin Minsky and Frank Rosenblatt (classmates from the same Bronx High School of Science). The debates can be articulated in one fell swoop—namely, (i) whether AI and Cognitive Science (which would lead to deep learning, and our current Chat-GPT) should try to emulate the actual inner neurological architecture of the human brain, whereby 'human *learning*' arises from a *singular mode* of neuronal binary/**digital** activity, (the nature of which is heavily reliant upon brute-force notions such as *locality, frequency* and *weighted strengths)*, or (ii) whether the brain's architecture—as was then and still is today so impervious to our complete understanding—should be modeled not based on its poorly understood neuronal architecture, but rather modeled on its computational performance and outcomes for such capacities as logic, *reasoning, cause-and-effect*. These latter processes are uniquely human and seem rather **analog** in nature, as they give rise to symbolic rule-based procedures of language and 'human *understanding*'. The Singular vs Dual Mechanism Model debates are currently ongoing in the field. These papers amount to some of my thoughts on the topic. The following links are pulled from *informal working papers and squibs and* represent some of my thoughts are the current state of a potential AI-to-Natural Language Interface. The last three papers (Section III), particularly 'Why Move?', attempts to capture this AI to Natural Language interface regarding developmental stages of child syntax.

This informal e-book is organized into three sections: Section I 'The Neuro Basis for Language' (presented herein), Section II 'Recursive Grammars', and Section III 'Child Language Acquisition' (links).

Contents

Section I: The Neuro Basis for Language.

3. Squibbing against continuity in AI

The fact that the brain is made up of neurons doesn't tell us much about the underlying representational mode upon which human thought is delivered, nor does it account for whether there are analogs to computer-software procedures as found in Artificial Intelligence (AI). The arguments herein contrast two types of neuronal delivery systems (local v distant, serial v parallel) in determining how short-term memory (hippocampus) tethers to 'local-domain' connectionist models, while long-term memory (cortex) tethers to 'distant-domain' symbolic models: thus, any putative interface which seeks to model the human global thought-process must require a hybrid model. The dual distinction, while model-based on serial v parallel neuronal processing, may provide insights into human language and cognition—for instance, we now know that Cortico-Hippocampal interplay (distant-to-local) shapes representational context in the brain. Hippocampal-Neural-net models (such a connectionist multilayer-perceptrons) seem to play an important role in the 'correlation' of local, frequency-based representations ('words')—whereby such 1-1 correlations can be readily captured by statistics—while Cortico-Symbol-manipulation is crucial to a deeper 'understanding' in spawning the necessary distant and recursive implementation which defines human language ('rules'). Another way to juxtapose these two distinct systems is to speak about the role 'Items' vs 'Categories' play in human language and thought—the former Item being advanced by brute-force statistics which promote 'local domain' correlations, while the latter Categories promote 'distant-domain' understand—such as logical inferences, causal relations and abstract knowledge. We believe the human mind to be uniquely defined by the latter categorical manner—viz., human thought is representational in nature, abstract in variable usage, and hierarchically recursive. We certainly know that much more goes on beneath what meets the eye in human understanding: broad understanding is certainly much more than the sum of its narrow parts. Any well-designed AI wishing to simulate human thought must capture these unique prerequisites.

https://www.academia.edu/103248724/Squibbing_Against_Continuity_Claims_in_Artificial_Intelligence_Why_We_Cant_Get_There_From_Here_The_Pursuit_of_Recursive_Neurons

Section II: Recursive Grammars

4. ABABABA-Grammars.

Recursive embedding as part of the language faculty has recently become the one essential ingredient in establishing the definition of what constitutes 'human language'—namely, recursion: that quintessential phenomenon which separates animal communication from human language, stage-1 child utterances from full adult syntax, MERGE operations over MOVE, and human abstract rules found in the human mind vs Deep-Learning/AI algorithms: Why child stage1 cannot discriminate between the expressions 'boat-house' vs 'house boat' (the former a kind of house, the latter a kind of boat); Why regular rule formations such as the prosaic plural {s} as found in English remain productive over an array of nonce (never-heard-before) items, and why irregulars must rather be memorized (via reinforcement: Stimulus & Response); Why Move operations provoke a non-frequency-driven recursion of the [[ ]] type, while Merge relies on frequency of item for brute memorization. Some argue that recursion is a recently adapted byproduct of a newly emergent human brain, perhaps having arisen as recently as 40KYA (thousand years ago), and perhaps the one feature which gave Cro-magnum an edge-up (in software) over Neanderthal. The following paper examines some basic issues surrounding the theme 'Recurrent vs Recursive' within a maturational/developmental progression: viz., syntax, long-distant dependency, and recursive design whereby algorithmic computation {x+y=z} governs grammar, as opposed to frequency-driven adjacency (x=x) as found in platforms of artificial intelligence.

https://www.academia.edu/108466141/The_Recursive_Linguistic_Mind_Recurrent_ABABABA_Grammars_Recursion_and_a_Note_on_Child_Syntax_1

5. A Note on Artificial Intelligence and the critical recursive implementation: The lagging problem of 'background knowledge'

Most historians of the Cognitive Revolution consider the now historic 1956 MIT IRE Conference 'Transactions on Information Theory' to be the conceptual origin of the revolution. It was at this conference that three of the most important papers in the emerging field of AI would be read:

(i) George Miller's Human memory and the storage of information (coupled with an earlier 1955 paper *The magic number seven, plus or minus two: Some limits on our capacity for processing information*).

(ii) Allen Newell & Herbert Simon's paper *The logic Theory Machine: A complex Information processing system.*

(iii) Noam Chomsky's paper *Three models for the description of language*. But it would not be long before splits would occur in the very defining of AI. For some, let's call them the AI-soft crowd, despite the ever-growing consensus that the brain really did not function like a computer after all, (as was earlier suggested by the naïve 'brain is computer' metaphor of the time), the AI-soft crowd, against the push-back, were content to go their own way and see just how far they could actually push their learning algorithms in solving 'real-world' problems (eventually using Bayesian networks). Most early cognitive scientists of this time—while now at least partially acknowledging and accepting the fact that what they were doing was indeed not real 'human intelligence' modeling—would nonetheless remain undeterred from learning about how to improve upon these non-human-like networks. One AI-soft champion that stands out here would be Frank Rosenblatt and his Perceptron model for visual learning (1959-1962).
https://www.academia.edu/104606120/A_Note_on_Artificial_Intelligence

Section III. Child Language Acquisition: The Maturation of Recursive Structures.

6. **Why Move?**

One of the leading questions burning in the minds of most developmental linguists is: To what extent do biological factors such as brain maturation play a role in the early stages of syntactic development? The theoretical framework 'Merge over Move' is applied here to the earliest observable stages of child syntax, a stage-1 which demonstrates a complete absence of movement operations owing to a complete lack of INFLectional morphology. The data in the paper support claims for a Non-INFLectional stage-1.

https://www.academia.edu/108466152/Why_Move_Preliminary_Thoughts_and_Overview_How_Merge_over_Move_informs_Early_Child_Syntax

7. **Remarks on a Minimalist Approach to Early Child Syntax.**

The working hypothesis in this paper is that the young children's syntactic parsers are initially unable to advance (MOVE) a morpho-syntactic utterance, both at PF (phonology form) and at LF (logical form) up the syntactic tree (whereby MOVEment would thus save the derivation from being sent off immediately to early semantic transfer). A pervasive deficiency of recursive movement is not just a surface-level PF deficit, but is also found at interpretation. Hence, as a metaphor for this lack of movement (both at PF and LF), children's early utterances are indeed semantically frozen deep within the prosaic trappings of the bottom portion of the tree (namely, within the base-generated VP phrase) and are thus sent immediately to spell-out. In this paper, I propose an initial 'merge-only' stage of child syntax which can account for a rather wide spectrum of implications leading to the impoverished state of early child syntax. Using Chomsky's current Minimalist Program (MP) framework, I adopt a 'Merge over Move' hypothesis as a developmental sequence thus accounting for the cited mixed word order, lack of inflection, and misreading of syntactic compounds found in the data.

(PDF) Remarks on a Minimalist Approach to Early Child Syntax | joseph galasso - Academia.edu

8. **Oxford Bibliographies: 'The Acquisition of Possessives'**

Developmental Pragmatics - Linguistics - Oxford Bibliographies

1

The Basal Ganglia, Astrocyte-$\mathbf{Ca^{2+}}$ Neuronal Circuit, Artificial Intelligence and the Dual Mechanism Model.

Real Requirements towards AI-Transformer-to-Natural-Language Interface.

The most compelling evidence to date for involvement of the Basal Ganglia (BG) (Basal Ganglia[1] Grammar) [13, 16, 20, 27, 48] in natural language comes to us from theoretical *movement operations* (nested dependency, distant binding and trace-theory). This implication of BG overlaps with well-established evidence showing Broca's involvement with movement [39]. Dual pathways are a marked characteristic of BG insofar that in cascading down-stream neural networks, both direct as well as indirect paths affect admixed neuronal populations from multiple cortical areas [20]. A tentative proposal may suggest that any notion of duality at the subcortical level may have the ability to simulate what we know of local vs distant binding dependencies as found in Dual Mechanism Model accounts of natural language [2, 7]. A theoretical (meta)-synthesis which seeks to connect what we know of Natural Language (NL) with current trends in AI/Transformers may offer us a potential merging of what has up until now been two quite disparate underlying systems. If we assume that NL systems mirror what we find in Parallel Distributed Processing (PDP) across neural networks [20]—and via extension be applicable to any putative AI/Transformer-to-NL corollary [47]—then, by definition, some component of the PDP would necessarily entail a *capacity-state* which corresponds to *concepts, symbols and categorial rules*—i.e., *real recursive-based* prerequisites for natural language which up until now have been sidelined in the implementation of AI modeling: such symbolic/categorial rule formation transcends mere itemized-style connectionism (typically predominate in past PDP-connectionist models [50]). The question put here—given the recently discovered properties of 'non-linear' neurons and neural networks—is whether an AI/neurological model can be envisioned which incorporates said recursive properties found in NL.

Moving beyond larger cortical areas to subcortical structures, many of which involve neuronal-modulate distributions at the synaptic level, this brief note—merely speculative at best, and synthesizing some of my recent writings on the topic [11, 14]—attempts to tease out what might be considered *real requirements* towards any putative *AI Transformer-to-Natural Language Interface.* It has long been sought out just how a dual mechanism model (dual-pathway) gives rise to natural language—viz., how the brain operates in *parallel* on gradient time scales whereby language input gets partitioned into two fundamentally different areas—relegating fast time-scales to *itemized learning*, and slower time-scales to *categorial procedures* [1, 19], (representing a morphological *dual-pathway* distinction associated with temporal vs frontal lobe partitions)[2]. (For contrast, recall that in traditional Recurrent Neural Networks (RNNs), input training is

[1] Andrea Moro (pc) suggests that it may be specifically the ventral dorsal head of the caudate which makes-up the largest contribution to movement as found in natural language. (Paper Draft: Nov. 3, 2023).

[2] While reaction-times studies in language certainly show fast-to-slow response times—as in the N400 millisecond for lexical semantics (Items: Wernicke's area) vs the slower P600ms response to grammatical anomalies for

done *serially* via linear-sequential networks). The parallel time-scale distinction may be understood in neurological terms and extended to artificial neural networks (ANNs) whereby some components of stimuli show sensitivity to frequency-effects (which peg to Bayesian statistical modeling), while other stimuli remain insensitive to such probabilistic outcomes (e.g., 'token-item vs category' distinctions as found in natural language). In generative linguistic circles, talk of such a cerebral *Dual Mechanism Model* (DMM) [1,2] has largely been used to support maturational theories of child language acquisition [3] as well as a plethora of studies dealing with Second Language processing errors (known as 'shallow-processing' [4]), Specific Language Impairment and Autism. As evidenced by fMRI studies, the human brain indeed does seem to operate in *parallel* (over gradient time scales), first locally, and then over nested hierarchical domains [5].

Backwards-engineering these facts, we can see that such brain-imaging results support recent notions espoused in theoretical linguistics, where fast-signaling local/MERGE-operations (itemized learning) distinguish themselves from slower-signaling distant/MOVE operations (corresponding to manipulation of symbols and rules [6,7,8,20]). If a DMM does arise in the human brain, one question to ask is how to extend the analogy to ANNs, such that a dual partition can become realized at the sub-cortical neuronal-synaptic level—a level up until very recently thought to be quite homogeneous and serial in nature? Some recent research into what has become known as the 'astrocyte-neuronal circuity' suggests such a possible DMM at the neurological/subcortical level [21]. The hypothesis is that 'slow vs fast' *wave oscillations* (so-called *cortical volleys*) related to $\mathbf{Ca^{2+}}$calcium-imaging might be one potential chemical neuro-transmitter capable of triggering such dual neuronal circuitry, suggesting a DMM at the neurological level. In an attempt to extend and tentatively match human brain-imaging to responses in ANNs, it's been reported that slower moving and perhaps long-distance single-neuron Ca2+ signals (action times: hundreds of milliseconds to tens of seconds) mimic what we see in slower non-linear/embedded recursive neuronal networks (RvNN) [11, 15] (see Galasso's Squib [11] for a review of Marr's work [15] on recursive neurons). This is opposed to what we find regarding the faster time-scaling of RNNs. It is believed that this slower time-scaling, which recruit spatially distant and diversly disbursed neuro populations, is ultimately what leads to the unique nested coding which gives rise to natural language/syntax—where long-term memory storage and subsequent short-term retrieval (episodic rehearsals) activate cascading neurons across divergent networks [9,10]. These kinds of episode rehearsals generate what has been termed Dense Associated Memory, providing the neuronal network with feedback loops thus generating recursive means [21].

The following note represents a meta-synthesis in attempting to merge together what we now know of (i) the Recursive Hierarchical Implementation (RHI) (in neuro terms so-called Recursive Neural Networks (RvNNs)) with (ii) Cortical-Glial substrates in the brain. The idea here suggests that there is a dual mechanism model of sorts in the way of astrocyte-modulations which peg to fast/short vs slow/long-wave modulations in synaptic oscillation— the former being pegged to fast, reflex-like responses [46] processed in the hippocampus, while the latter pegged to slower wave-oscillation signatures found in the cortex. The 'hippocampus vs cortex' functional distinction has been long held in neuroscience and may be viewed here to help shape a dual capacity of neuromodulation, which in turn can lead to recurrent vs recursive

functional-abstract grammar (Category: Broca's area)—this somewhat simplistic Wernicke vs Broca area split perhaps only addresses the larger cortical areas governing language. In recent research, more fine-grained analyses reveal that more of the action might be taking place at the sub-cortical neuronal levels (as this paper attempts to show), with Astrocytes Glia cells being perhaps the crucial component related to a putative dual-path synaptic interaction.

neuronal implementation in the brain. One way to tease out such dual capacities within ANNs is to similarly evoke classic distinctions between Old-Fashioned Artificial Neural Networks (OFANNs) and very recent discoveries in RvNN Transformers (e.g., ChatGPT-3/4, Bard).[3]

It goes without saying that human-brain computation is of an emergent nature, is of a non-linear and non-sequential order, is recursive rather than recurrent in architectural design, and, perhaps above all, is defiant to all norms of statistical learning. Such computation assumes a level of abstraction which goes beyond mere levels of external input, thus creating internal models of the outside world by means of subtle abstraction—so called Theory of Mind (ToM) procedures. These abstractions proceed in stepwise-cyclical fashion by the combining and (re)re-combining (recursively) of rules, symbols, and categories to the

representational mapping onto real-world token items and events [15]. The coding of such representational mapping must include the usage of algorithmic variables (a move away from item towards category). The notion of category is of interest to us since it suggests that variables are being read off by other variables in a recursive fashion—and that such 'self-attention' (bootstrapped by *embedded theories)* may play a role in higher-order cortical processes such as learning, long-term memory, and language. The aim of computational system-science is to somehow notate these cyclical processes which take place in the brain in algorithmic/mathematical terms. Neuro Computation may not be restricted to only include feed-forward, and unidirectional flows, as attested in OFANNs, but may also support the kind of internal feedback loops which allow subtle adjustments, indicating the kind of self-attention thus far unseen in typical top-down RNNs. Linear Recurrent models seem unable to support the type of self-attention indicated by these newly discovered Transformers. The question now becomes how can we implement the RvNN mechanism biologically? Might there be any way to accommodate this dual capacity in a biology which otherwise suggests a neurology to be homogeneous in nature.[4] The workaround here would be to posit, at the very minimum, at least two distinct modes of Neural-Glial function which display such a putative dual makeup. Recent studies suggest that indeed a dual pathway of Neuro-Glial-Circuit connectivity does exist (presumably at the Basal Ganglia-Caudate region [16], fn.1).

And while exactly how the brain works currently eludes us, we do know *how the brain doesn't work* [40]. The question confronting those of us working within Artificial Intelligence (AI) (neurocognitive scientist and linguists alike) is to ask how such non-linear human brain capacity can become represented within current AI frameworks. While how the brain works is still not well understood, there is an

[3] Some cognitive neuroscientists take the view that Transformers can be built exclusively reliant upon recurrent operations. If so, then what we might add to that claim is that the recursive implementation is therefore somehow part of the inherent design of the platform (perhaps hidden and part of the architecture). In other words, while recurrent models may seemingly simulate much of what is behind a full-fledge GPT-Transformer, it is our view that the necessary coding behind such high-level (natural-language like) GPT operations must be recursive in nature. While just how the recursive AnBn algorithm is implemented in the platform may be of open debate, as the matter stands, a mere RNN networks fails to have generative properties to meet what would be the real requirements behind any natural language model. Recursive A^nB^n yields embedded/hierarchical nesting [a [a [a b] b] b] (distant bindings)… while Recurrent $(AB)^n$ yields so-called linear/non-hierarchical 'ABABABA'-grammars [AB], [ABAB], [ABABAB] hence, [ABABAB(A)]-Grammars ([13]. The latter is strictly dependent on brute-force BIG DATA-calculations and probabilities, a dependency otherwise not based on structure.

[4] Despite acknowledgement of heterogeneous cortical functionality, it was always assumed that the make-up and function of sub-cortical neurons and neuronal networks were homogeneous in design and function.

overwhelming consensus that mere 'input-to-output' strength-weighted **[ ]–[ ]** modeling (feed-forward or hidden-layer connectionism) falls well short of capturing human-like thought. In fact, all such recurrent top-down trained devices (current as well as bygone artificial neural connective networks) simply cannot capture human thought: not only are their performances dismal, they are catastrophic! [29]. What is now well appreciated is that human language is of a quite exotic 'recursive-embedded nature'—viz, a 'many to one' // 'one to many' **[[ ]]** unfolding algorithm whereby manifold of overlapping matrixes layer upon a singular event potential, and that such overlaps may in fact recursively generate a kind of 'shadowing' effect whereby the 'brightening and dimming' of any given neuro-stimulus complexity becomes part of the neural network. It is the promise of a new kind of neural network, referred to herein as Transformer, which we speak of in testing whether such modeling is brain-like in nature and is given to human-like performance. This 'ebb and flow' (brightening and dimming) of the on-line holding of memory during decision-making tasks [17], and memory rehearsal and replay during memory consolidation [18, 11] seems to activate the kind of neuronal distinctions which cut across astrocyte-wave oscillations. (Nb. What seems to be a unique feature of human thought processes is this interplay between working and long-term memory [44]. In the *declarative/procedural* model [1], an overlap of such processing distinctions may play out in what we find in natural language between lexical words (items) vs syntax (category), respectively.

In order for an AI-Language Transformer to perform similarly, its processing must have simultaneous access to both a (vertical stacking) of a list of words (a lexicon) as well as a list of sentence strings stored (a syntax). One way to do this might be to treat all sentences as unsegmented lexical chunks (i.e., as lexical items, *per se*). But this gains us nothing beyond mere item-based learning. So, while the devil is in the details regarding the real architecture behind CHATGPT/Transformer transcription, it minimally must have a dual capacity to process (i) items on one hand and (ii) a symbolic syntax on the other (See Marcus for review [28]).

This short synthesis highlights what I think are the REAL requirements behind any putative Transformer-to-Human Language (THL) interface (e.g., ChatGPT-4, Bard, or other bio-implementations towards AI). The five requirements outlined below do not represent an exhaustive list but are presented merely as pointers to what many in the cognitive sciences believe to be real and necessary prerequisites to a viable THL interface. The one analogy I provide below is that of '*Pentimento*'[5] (a kind of modulating dimmer-effect between *brightening and dimming* which creates 'shadow' overlaps which mediate between surface (short-oscillating) and deep (long-oscillating) networks. This oscillating factor is, I believe, quite essential in determining the putative Ca2 wave fluxes correlating to (i) short-term memory (hippocampus) vs (ii) long-term memory (cortex): the latter now seemingly being related to so-called **Astrocyte** Glia cells, now widely discussed in the literature [19-22]. The idea currently being pursued is that such Astrocyte three-prong formations—*three-prong* since they don't constitute the traditional 1-1 neuron-to-axion nexus, but rather lay down an overcoat sheath (similar to myelination)—undergird relevant neuro networks responsible for human language[6].

[5] The notion of Pentimento is based on a Lellian Hellman poem 'Pentimento' where surface imagery my overlap and superimpose onto hidden deep-structured layers.

[6] The magic number 3: it is well known in physics that measurements are complicated when three or more variables enter into measurement/calculations (similar to Heisenberg's *uncertain principle)*. Similarly, the unique 3-pronged

Regarding this 'dimming effect', this would have been referred to in early connective networks as so-called 'strengthening & weakening' of weighted values. Such previously held (top-down) Artificial Intelligence (AI) networks—which tried to seek bio-relevance, originating as early as the 1940's Hebbian school of 'What fires together wires together'—has been very recently supplanted by bottom-up transformers which, being trained on a very minimum innate architecture, become not only *self-learning*, but also *self-aware* (i.e., self-attentive). Cognitive scientists like to say that Transformers are 'mysterious' in this way and rather opaque, like the human brain/mind, precisely because transformers are based on an untrained bottom-up platform. On the other hand, Old-Fashioned Artificial Neural Networks (OFANNs) are not so mysterious as their modeling is quite transparent based on the human input trainer.

Circuit Dynamics.

Some circuit dynamics which fall within the range of <millisecond> (ms) seem to control fast-reflexive and compulsive behaviors, tethered to short-term memory (e.g., perception and decision making, lexical selectivity of vocabular [23, 46]). Slower-moving synaptic changes <high-length ms range: up to lasting minutes and hours> may control long-term memory, learning and language [24]. It is this latter slow-wave which is able to seek new information against the backdrop of previously constructed embedded models [20]—the core of such embedded operations seems to be responsible for selective attention and language.

The dynamics of Connectivity suggest two main patterns (RNN vs. RvNN, respectively): (i) linear feedforward, supporting a unidirectional flow of sequential information, and (ii) recurrent, composed of positive and negative feedback loops that lead to self-sustained multiple activity patterns [22]. It is the second, recurrent dynamic which we believed, if heightened in response, may move beyond being a mere statistical tool, and rather lead to substrate codes in the brain capable of handling variables and algorithms (i.e., a recursive-hierarchical implementation in the brain). One way to look at this is to say that linear sequential feed-forward flows may indeed become superimposed (onto itself, e.g., degerming self-attention) if the right kind of neuronal circuitry is employed. The idea here is that it is the unique features of astrocytes (a three-pong glial sheath) which enable the circuitry of RNN to backslide onto itself, which in turn leads to self-attention. By self-attention, we mean the ability to read not only the input bits coded in the representation, but also the source of coded material itself, triggering a recursive A^nB^n/RvNN computational signature [13]—namely, an RvNN operating scheme at the neuro-substrate level is necessitated by the evaluation of new information against the backdrop of previously constructed embedded information, hence a nested construct.

In the past, it has been proposed that such evaluation is probabilistic in nature [25, 26]; however, rethinking such shortcoming based on Bayesian models has led to a reanalysis of the necessary recursive implementation. Surely, natural language goes well beyond Bayesian statistical operations [43]—hence, any such crude statistical tools such as generalized linear models (GLMs) deemed worthy of governing Transformer and/or Natural Language most certainly would fly in the face of what we know of natural

glial astrocyte functions very differently from typical neuronal bidirectional heads as they seem to provide feedback loops necessary for approximate values/variable over brute-force statistics [15]. This may give the computational function an added feature having to do with category/variable functions.

language (Chomsky). It is worth noting here that Chomsky often talks about the brain/language as a 'black box', even a 'ghost in the machine': we really don't know what is going on inside (the brain), how it works. This is because natural language is an unsupervised 'bottom-up' phenomenon: okey! to a certain extent 'top-down' if you consider the Faculty of Language (FL) to be 'Innately Design', but bottom-up in terms of a self-learning procedure. There is no top-down, language-specific guide to supervise the child in her language acquisition, and even if there were, we know that children go well beyond their supervised data, while at the same time never producing so-called unconstrained 'wild' grammars [27]. It goes without saying that there is no preselected labelling of language beyond the basic instruction: MERGE items {α,β}, with local vs distant MOVE options [6]. Well, the same analogy recently seems to be employed when talking about ChatGPT: We say things like ChatGPT is 'mysterious' (a black box) and that we don't really know how it works. We claim we don't know how the transformer goes about learning or correcting itself. Well, this mystery can only be maintained because there is no top-down 'software engineer' behind the scenes guiding the selection of choices to be considered. There is only the bottom-up accumulation of BIG DATA. (Nb. Though I think we do carry this analogy too far: we really DO know how GPT works—that is, if it is a model solely reliant on (OFANN) probabilities (which still seems to be the operating platform of choice these days [28, 29]).

If, on the other hand, GPT is truly a recursive design operating on variables, analogous to nature language (child language acquisition), then I accept the premise. I must point out here that in the past, GPT modelers have been extremely weary of using true rule-like symbolic procedures (such as symbolic tree diagrams) precisely because of the sticky problems imposed by an operating network reliant upon the tree diagram (top-down) as the source of linguistic data[7]. In other words, the tree-diagram itself introduces

inductive biases (i.e., the assumption that the data must flow from the tree diagram, (rather than the other way around)). Recall, that in linguistics, the tree is simply a notational device (albeit physiologically real in notation); the data a tree might provide might only represent a surface-phonological string (the upper spell-out, or left periphery post movement). We know that children go well beyond the data in consistent ways. If assumptions (biases) dictate that the [tree = data], say on a 1-to-1 sound to meaning, then the model will surely fall short: the dilemma is rather than acquiring new representations bottom-up, GPT modeling may simply reuse a preselected tree diagram in order to recombine the same old structure. We know that surface strings while seemingly similar in sound structure (phonology) may take on very different deep structures (whereby semantics/pragmatics are required for interpretation). In other words, a wrongly selected inductive-bias assumption which directs the data to follow a specifically pre-selected tree hierarchy (top-down) may cause a failure at interpretation (and thus prevent the learning mechanism from expanding). This is just one such problem. The best example of this is the classic textbook

[7] One great advantage of symbolic systems is that they are not overly depended on data size: they can work over an array of low latent variables across nested hierarchy. OFANNs and other strictly recurrent ANNs solely rely on BIGDATA in order to meet probabilistic tendencies (See [30], 'Sentence no. 4' for an instructive account of how language is not 'data dependent' (= *count every example*), but rather is 'structure dependent', a more abstract property (= *every example counts*).

sentence: 'John saw Mary with a telescope': while having the same surface phonology, only one of two possible trees gives the right interpretation that it is 'Mary' (and not 'John') who is holding the telescope.

In today's Chomskyan 'Minimalist Program' terms, Merge of {α,β} [6, 7] (leading to phrase projection) is an exclusively bottom-up enterprise (though guided by top-down semantic-pragmatic sources). Such a fountain effect of top-down (Language Faculty) but bottom-up merge-based phrase projection may certainly wreak havoc for any OFANN recurrent operation. So, if the GPT operating algorithm believes it is the actual tree (top-down) which is the computation source (inductive bias), then we are in trouble. Symbolic tree diagrams are still not commonly employed in ANNs for this reason— symbols are hard to work with, not as robust as recurrent Bayesian models, and may be better suited for specific tasks (so-called expert systems).

Neural Substrates: Brain Oscillations.

The implication here is that the brain codes/oscillates (at the neural substate) at multi-levels: (i) at fast/linear levels of encoding of Evoked Action Potentials (EAP), and (ii) at slow/non-linear potentials—whereby the former could be seen as a strictly recurrent operation (RNN equivalent), and the latter a potentially recursive operation (RvNN equivalent). This means that different arrival-onsets of stimuli within millisecond intervals can impact individual firings of neurons (see (1) below). These arrival-onsets are determined by Ca2+ transients. What this means is that a singular bidirectional neuron can take on varies dynamic roles as determined by these transients. What has been uncovered is that fast reflex decision-making tasks can be as ultrafast as 10ms (falling within the range of <100ms), while that same neuron can adapt to encoding slower astrocyte wave-functions (up to minutes) when deliberation occurs (with language and higher cortical reasoning falling on the slow side, as would be expected of RvNN delivery systems). Recall, that grammatical/recursive anomalies based on recursive-systems errors peak at around 600ms (P600) [31]. Further still, we now know that short-term memory time-scales incur fast-wave fluxes (hippocampus related/RNN), while long-term memory incurs longer-spatial wave fluxes (cortex related/RvNN). The ebb and flow of pulling long-term memory up to short-term working memory (as in a temporary file for a computer) and then pushing the same short-term code back to long-term residence (as the default) constantly rearranges the actual neuronal infrastructure such that new neurological connections/rewiring have to be made [see Squib [11] for review). Given that fast time-fluxes involved with working memory have to be constantly reset (to the default slower-cortical signature), the persistent firing of ebb and flow neurological activity achieves 'exquisite tuning' of recurrent circuits [32].

Extending the analogy of probabilistic Hebbian modeling (OFANNs) to fast-wave lower-level cognitive domains, we can begin to tease out findings which show that category and variable coding (presumably at the Basal Ganglia-Caudate regions of the brain) becomes selected whenever higher-brain functions are call upon. **Ca^{2+}** (a chemical neurotransmitter) is the transient marker, with slow-wave oscillations serving higher functioning (cortex) canonical operations [33]. The marker may encode additional, external feedback-(loop) variable information to the dynamic circuitry (synergistic and/or complementary in nature [34]. Neuromodulations refer to this ebb and flow of rapid to slow oscillations of circuits whereby any reconfiguration of such chemical neurotransmitters (acetylcholine, dopamine, noradrenaline, and serotonin) can reconfigure the electrical releases of the subcortical and brainstem nuclei [19]. It has been assumed that fast **Ca^{2+}** fluctuations surrounding 'local-domain' circuitries are too anatomically fixed to be utilized for more distant, and perhaps 'global-domain' circuits, whereas slow-wave **Ca^{2+}** correspond to highly complex and content-dependent variables and may continue to have down-stream effects on other non-local domain circuitry. This same 'local vs distant' distinction is similarly a feature which shows up in natural language theories (Chomsky 1995 [6]), whereby the bottom-up phrase-projection of local/linear Merge {α,β} is then extended (superadded by extension) to a distant/non-linear MOVE operation {α {α,β}}, showing recursive hierarchy [7].

Astrocytes. Astrocytes—behaving as switches (Janus-headed gatekeepers of neurons)—have the benefit (a unique property) of not being too/overly sensitive to electrical excitable neurons. Given that their release is chemical in nature, slow percolation of firings coupled with potential cascading downstream effects can give rise to more global computations (e.g., variables & algorithms) across distant neural networks. An added feature is that astrocytes are not neurons *per se* (but rather Glia formations which serve as a sheath-like covering to neurons). This allows them to interact with a bidirectional neuron/axion in manifold ways. The nature of their three-prong tripartite-head may allow continuous feedback loops to enter into the otherwise bidirectionally-fed neuron. The unique design/function of astrocytes suggests that they may generate so-called logic gates, the kind of formal, canonical operations which may govern natural language (e.g., AND, OR, NOT, XOR, NAND-operations) [35][8]. Other factors may play a role in astrocyte global/canonical computations—viz., other interactions involving neuro-messenger molecules (such as Na+, IP3, cAMP, which work alongside Ca2+ in creating ion-based signals [36]. Such a dynamic astrocyte system has been unpacked in recent research showing an underwriting of 'movement-based' and (potential) recursive implementations, as would be found in language [37]. On this higher end of the dynamic spectrum, non-linear ANNs are increasingly being used to replace old-fashion linear (OFANNs) for machine learning. However, it is our belief that in order for non-linear ANNs to truly live up to their name— to be generative models, to manipulate low numbers of latent variable, and not be simply turbo-charged Bayesian models—their underling computational systems must approach recursive implementation, or, at the very least, be a system which utilizes both (local) linear statistics alongside (global) hierarchal symbolic 'tree-like' structures, thus promoting *hybrid systems* which simulate Dual Mechanism Models a found implemented in natural language ([3,28,29]. As far as 'natural language' goes, we know that it is the frontal cortex which is responsible for man's ability to create internal cognitive maps and representations of our external worlds. It is the frontal lobe which diverts token items (+Freq(uency) sensitive) into -Freq categories. A buffer-zone of sorts would need to be in place to act as an interface. With this kind of ongoing work into neurological equivalence, it goes without saying that the exact nature of this astrocyte-Ca2+ interface, I believe, is quickly becoming the holy grail in understanding not only the dynamics behind machine-learning (and the like) but will also inform a more complete understanding of lies beneath natural language.

Self-Attention. It has been reported [38] that slow-wave modulations (a default mode of cortical higher function) can provide the kind of feed-back loops which can recruit self-attention of neuronal firing. In other words, this self-attention allows the evaluation of new information against the backdrop of previously constructed embedded information (a kind of mirror onto itself: perhaps not unlike what we know of actual 'mirror neurons'). It is speculated that such stimulating 'ebb and flow' between (i)

[8] I would like to consider 'OF' also as a canonical operation (recursive in nature where nested expression are looked-up and 'negated' upon <NOT>, as in the expression: [an enemy [OF an enemy [is a friend]]].

feedforward down-stream effects (RNNs) followed by +/- feed-backs loops (RvNNs) lead to this kind of self-attention. Of course, there are still cognitive-science holdouts that ANN (even OFANNs recurrent RNNs) might be in a position to similarly give rise to self-attending processes. However, it is our claim that any self-attention would have to be embedded in the kind of dual processing which utilizes as one of its two forms a proper recursive operation.

Five Requirements (minimally) for AI-to-Natural Language Interface:

1. *Graceful Degradation*. Where gradient even cascading errors do not necessarily bring a system to terminal status (stoppage). While catastrophic failures (total malfunctions) are a typical feature of top-down (supervised) systems, (e.g., linear-sequential Recurrent Neural Networks (RNNs)), bottom-up self-attending RvNNs tend to allow processing to approximate in light of systems errors and noise. Humans' capacity to think on one's feet (using background knowledge, experience, content, even intuition and guesswork) are all hallmarks of the abstract/high cognitive levels of human thought.

   Top-down vs Bottom-up Processing: The ability to incorporate what has been learned in its prior knowledge. This incorporation happens at two different and distinct levels of processing and speaks to a unique ability to approximate information from the two networks—viz., where bottom-up represents environmental (bits of info) as delivered to the closest surface-interface system, but where top-down context (prior knowledge, inference) may approximate otherwise direct 1-1 bits to a 1-many scattering, putting environmental stimuli in check. The result of the two reaches a final state, referred to as a *constellation state*, whereby attractor bundle of features become calculated via *strengths* and eventually settle on an appropriate assembly of linking networks (the so-called **binding problem**). Here's a way to think about the two processes. Let's take a simple *visual cortex* scenario and apply a binding of 'item to category'- constellation: suppose you see two items: a *car* waiting in front of a *stop sign.* Ok, you say: there's two bits of neuro inputs (item) which can easily be accessed together via association rendering a 'traffic scenario' (category). The ITEM/OBJECT [CAR] fires its appropriate high-level/categorial networks of thematic and taxonomic neurons & [STOPSIGN] fires its appropriate neurons. But now, say, a third bit of info enters into the binding (a *theatrical stage*) [STAGE]. This subsequent third bit of info (item) will force new modulations, enacting different levels of binding across the constellation-states imposed on by the prior first two firings (via hundreds of back-and-forth volleys across distant cortical regions). This amounts to a new approximation inferring that this visual field is no longer a (real) 'traffic scene'—viz., that the AGENT *car* is not in an ACTION state of being *driven,* so that the neuro scheme *[AGENT + (ACTION) + INSTRUMENT]* no longer gets processed since the INSTRUMENT *stop sign* cannot deliver any thematic value. This inference will produce a new subsequent settled state (with a new bundle of neuro wiring and firing)—viz., that this is a (fake) 'staged production'. (Note how the actual visual SEM/thematic items themselves don't change, but only the *intensity* of their bound constellation of their *attractor states)*. Such modulated constellation of top-down knowledge in this way is an accumulative result of this 'dimming & brightening', 'slowing & quickening' oscillations of neural networks, pulled from local and distant multiple cortical areas [49])—deriving context, inference, and pragmatic world-understanding. The functions of *basal ganglia cortico-thalamus* connectivity seem to play a critical role in pulling these diverse areas together in settling a constellation-state [20]). In this same context, see [29] (p. 23) for how a [STOPSIGN] graffitied with stickers generates such a level of (unrecoverable) bottom-up *noise* that a ‘google captioning system’ mistakenly identifies it as a ‘refrigerator door’ covered with ‘posty-notes’ (since brute-force probabilities as gathered from BIG DATA algorithms pull such [flat surface objects + cluttering of stickers] to neuro-associate with American-style refrigerators). In this sense, deep learning can be very *brittle*, and not robust enough to eliminate noise in the input. In sum, top-down is not merely the sum of its bottom-up parts, but rather ‘bootstraps’ itself to an entirely higher/different level (a level which may altogether be disconnected from the bottom-up source). Having a ‘deep-embedded theory’ about how the world ‘categorically’ works around us, independent for the day’s ‘itemized’ surroundings is such an epi-phenomenal bootstrap.

2. *Manifold shadowing*. The ability to create transparent-like overlaps (what I have called *pentimento effects*) at the subcortical-neuronal level. Presumably, such a dual mechanism model would have similar inner-workings to what we find in natural language (and may mimic what we find between recurrent v recursive processing distinctions). Such a dualling direct vs indirect cascading pathway model which affects/modulates admixed neural populations may play an essential role in any Basal Ganglia-Astrocyte interface system, further informing our understanding of the dualism behind natural language.

   For instance, in (1) above, while the items ‘car’ and ‘stop sign’ would bundle in local neurological domains such that the two items prime together, the third item ‘theatrical stage’ incurs no such local/bundled-priming and would rather trigger distant neuro-connectivity. When a distant neuronal triggering sits and overlaps onto local networks (pentimento), one can think of this as a kind of ‘shadowing’ (or superimposing of neurological connections). Such shadowing gives rise to neuro approximation, context and inference—all essential properties needed for any AI system to be deemed human-like in understanding. Part of this problem has been written about regarding *Binding*—viz., the capacity to link neural representations in the brain both a local as well as distant domains. Also, potential back-and-forth (volleys) between the two processing may further add to the complexity of thought [20].

3. *Short term vs long-term intermingling of synapses.* The Ca2+ fluxes seemingly are what’s behind the aforementioned neuronal dual pathway. With potential analogies between local vs distant volleys as stated in (2) above, this ‘ebb and flow’ of neural recall—recall to working memory/Hippocampus, followed then by reconsolidation back to long-term memory/Cortex—instigates a pruning and rewiring of neural circuitry to the extent that the oscillating patters ultimately seek out self-attentive/recursive processes (see *Dense Associated Memory*, providing the neuronal network with feedback loops thus generating recursive means [21]).

4. *Looping effects* (which require three-prong astrocyte glial cells). There is no question that perhaps the holy grail of human language has to do with such a recursive looping effect (3 above). Chomsky’s Faculty of Language (FL) [42]. Hauser, Chomsky and Fitch characterize this as FLn (Narrow) which perhaps only involves recursive syntax, with LFb (broad) being relegated to everything else (phonology, semantics, working memory, problem-solving). The ability for a system to look inside of itself via embedded theories, (say, at a deep-level, or even at its architectural hidden level) and gain additional information to percolate back to surface-level inputs may be what is behind this unique process termed ‘Dense Associative Memory’. Looping capacity has long been

sought after as the quintessential feature separating a system of *understanding* from that of a mere system of *knowledge*—the former being a horizontal rule-based/symbolic system, the latter being a vertical/associative push-down stacking system.

*Whereas in AI terms, FLb might be considered as operations reliant upon BIG DATA, FLn would be reliant on abstract and rule-based causal relations governed by recursive operations. Others suggest that the dual distinction can be viewed as network distinctions between auto-associator networks (FLn) vs pattern associators (FLb) [20].

5. *Sleep for consolidation* (a very bio-specific activity). As part of any biologically based *reconsolidation*, (a calming of back-and-forth volleys spread across cortico-thalamic networks) sleep becomes are overwhelming factor. I'd like to think that what makes us humans special is our capacity to sleep and reconsolidate the day's events into compartmentalized classification, too a self-attending matter at the cortical and perhaps even subcortical levels of the human brain. While dubious, there may some attempt to consider non-bio/AI analogues to sleep: computers may be capable of a 'kind of sleep' if they can parallel process both online and offline (IN SYNC), as humans do. (Again, I am personally very suspect of any real AI analog to human sleep—but the notion has been bandied-about in the literature). If indeed there's evidence of AI on/offline processing, I would count the offline part as a mode of sleep. The following is how one might go about assessing the notion of AI/computer sleep (perhaps some components of this can be instructive for us):

Note: Online AI-algorithms work with specific input/data. Such declarative/online algorithms improve incrementally as new info gets incorporated into the training/data-set. Stochastic systems with back-propagation, as used in RNNs, are a good example of traditional online sequential/incremental processing. Offline learning algorithms on the other hand work with data in bulk (i.e., all at once over an array of local and/or distant domains). Offline algorithms may often require secondary components such as generative models to support their online counterpart, enabling adjustments of the underlying algorithm. Such top-down offline learning algorithms need to be re-run from scratch at each take in order to (re)learn acquired data reallocated from previous data (and therefore may have the computational ability to infer casual relations between the two on/off-line processes). Similarly (though not necessarily as a result of off-line reasoning) humans' unique ability to formalize *causal relations* (cause-effect) has recently been found as being one of the few special properties required before any deep-learning system can be putatively claimed as having human-like reasoning skills [41, 43].

EndNote: Distant Binding from Linguistics to Neuro-computation and Marr's recursive neuron.

As an endnote, I'd like to make explicit how I think a basal ganglia (BG) grammar, coupled with astrocyte wave fluxes, might generate natural-language (NL) outcomes [45]. Firstly, let's review a few findings, tentative though they may be.

1. *Distant Binding* (anaphoric/antecedent coindexing, word/phrase-level displacement, nested structures) is a well-known feature of NL syntax, and theoretical distinctions on the SEMantic level can easily be made between (i) SEM/Local associations (attributed to lexical MERGE of {α, β} in forming phrases—e.g., as found in a verb phrase [VP drink milk] and (ii) non-local SEM constituencies

which may cut across phrasal boundaries (and reliant upon SYN). For example, consider how sentence (a) replete with SEM survives correct processing by Broca Aphasics (BA), while the same subjects suffer catastrophic syntax failure (b) (noting that BA subjects perform poorly on MOVE-based SYN [39]):

(a) [The apple [the boy ate] is red].
(a') What is red? BA response: 'the apple'! : (correct: [apple = red]/SEM

(b) [The girl [*the boy chased] is tall].
(b') Who is tall? BA response: *'the boy'. : (*incorrect: girl and/or boy = tall]/SYN

(Note: BA subjects default back to a flat-[ ] linear SEM processing devoid of any recursive SYN [[ ]]-embedding). (BA subjects process the closest item/subject of the modifier). (See [13]).

Normal functioning distant-binding would process as:

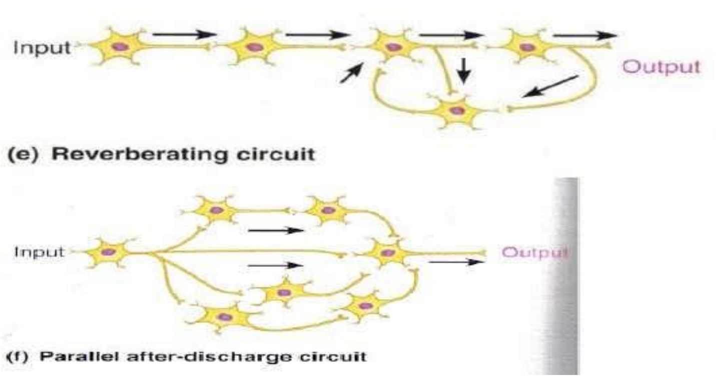
(b'') [The girl [the boy chased] is tall]. : (showing embedded hierarchy).

Neurologically, this can only be achieved by a recursive implementation in the brain. (See below for Marr's seminal work on the notion of recursive neurons).

Though the response in (a) seemingly shows correct embedded processing, a closer look reveals otherwise: the info in (a) is readily available based on local/surface-string info (semantics/pragmatics). This is not so the case for (b) which is ambiguous since both a 'girl' and 'boy' can be tall. Hence, in order to get at the correct response, embedded/SYN hierarchy must be employed. Several studies [7, 39, 48] indicate that BA have trouble when it comes to non-local & non-linear (distant) SYN-operations which rely on the ability to process embedded structures. In terms of an AI-transformer-to NL transcription, it seems universally that SEM features also encode fast and locally (as measured based on their EAP signatures).

In sum, regarding EAP distributions overall, fast-wave-related EAPs tie to recurrent SEM (RNNs) while slower-wave tie to recursive SYN (RvNNs). In child language, we see the same early distribution with recurrent/Thematic <AGENT VERB PATIENT> constructs emerging before more abstract/SYN constructs. Such neuromodulations—*back and forth cortical volleys* (between long-term and short-term memory involving of *insular cortex* [44])—have become an important factor in the ongoing research [20] and poses a problem as to what might be the 'relegating-factor' involved. Certainly, Hebbian neuro-to-neuron states make sense in providing the faster EAPs, but the question is how do these EAP signatures come to be represented at the subcortical level. One idea presented in much of the research cited herein suggests that it is the intervening astrocyte (perhaps buttressed by BG-Thalamus system [20]) which acts as a kind of buffer between SEM and SYN operations. If so, then there might be good reason to speculate that indeed a DMM becomes realized at the subcortical level, whereby lexical items (heavily frequency sensitive) take on sequential processing (as found in e.g., the verb phrase), while categorical SYN representation (reliant more so on abstract rules and less frequency sensitive) take on recursive cortico-cortical 'feedback loops which may bring together and bind divergent brain regions for consolidative processing. (Consider below). Given this feedback loop, AI/Transformers (as found with neuro circuity) can enjoy long-range dependencies between words and thematic-structured sentences.

Marr's Recursive Neurons. The late (and sorely missed) neuroscientist David Marr, I Believe, was the first to ever proposal that neuronal connections may take-on recursive-loop functions. The closest analog to recursiveness, at that time, happening at such micro-levels, probably related to what we thought was happening with mirror neurons, whereby interlocking bidirectional neuron firings might mirror each other in forming feedback loops. The notion that the transition loops of GFPs (gestalt frame potential) from (i) short-term to (ii) long-term, back to (iii) short-term memory retrievals, which indeed adds noise to the neuron, nonetheless may not necessarily mean that the neuron is completely eliminated, as was at that time sometimes suggested. Rather, it is currently thought that given a recursive neuron connection, amplifications may ripple through the neuron-bundles via dendrite-axion pathways in quite dynamic ways, allowing memory-traces (pentimento) of the original stimuli (first time-step GFP) to remain, while also incorporated additional overlapping information, (sometimes referred to as 'noise'), rendering embedded feedback loops. Marr went on further in suggesting that neural networks can exhibit pure recursive functions having to do with reverberating circuits. Marr suggested that while synapses become excited, information packages will pass along the chain of neurons, with the last neuron in the chain be amplified or attenuated. Marr stated: 'However, the output neuron also feeds back the same output message back to another neuron, which then loops the information back to the penultimate neuron in the chain'. In Reverberating or oscillating circuits, the incoming signal travels through a chain of neurons, each of which makes collateral synapses with neurons in a previous part of the pathway. It is this 'collateral' aspect which suggests the potential for recursive reverberation. Notice how input impulses run parallel, reaching the output cell at different GFP time-steps. In parallel processing, the input travels along several different pathways to be integrated in different GFPs timesteps and regions. It is believed that it is this parallel processing which allows human neurons to take on recursive functions. Marr had always speculated that it was the *pyrimidine neuron* which held the key to such recursive looping. Coupling this with what we now know today of astrocyte-basal ganglia-thalamus function, perhaps there is a path forward in bringing AI/Transformer systems even closer to natural language. While current state-of-the-art AI/Transformers may not be there yet [29]—and there are still plenty of cognitive scientists [28] and linguists [6] willing to bet the house that we will 'never get there from here' (and I count myself amongst them)—nonetheless, we seem to be approaching a *pivotal point* in realizing a possible convergence, bringing AI/Transformers on par with our understanding of the neuro-circuity behind a brain-to-language corollary. Below, we see Marr's original scheme for a recursive 'reverberating-circuit' neuron.



References:

[1] Ullman, M. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience* **volume 2**: 717–726 https://www.sciencedirect.com/science/article/abs/pii/B9780124077942000766?via%3Dihub

[2] Pinker, S. (1999). *Words and Rules*. Basic Books

[3] (a) Galasso, J. (2004). Towards a Dual Mechanism Model of Language Development. (Presented at Child Language Research Forum, Stanford). https://www.academia.edu/775453/Towards_a_Dual_Mechanism_Model_of_Language_Development
(b) ___Why Move? J. Article. *Open Science Framework (OSF)* https://osf.io/5he2x

[4] Clahsen, H. & C. Felser (2006). *Continuity and shallow structures in language processing*. CUP.

[5] Murray, J., A. Bernacchia, D. Freedman, R. Romo, J. Wallis, X, Cai, C. Padao-Scioppa, T. Pasternak, H. Seo, D. Lee, X-J. Wang (2015). Hierarchy of intrinsic timescale across primate cortex. *Nat Neurosci* **17**(12).

[6] Chomsky, N. (19950. *The Minimalist Program*. MIT Press. https://doi.org/10.7551/mitpress/9780262527347.003.0003 [7] Galasso, J. (2016). *From Merge to Move*. LINCOM Studies in theoretical Linguistics **59.**

[8] ___ (2021*). Reflections on Syntax.* Peter Lang Publications. DOI: 10.3726/b18267

[9] Vogelstein J., A. Packer, T. Machado, T. Sippy, B. Babadi, R. Yuste, & L. Paninski (2010). Fast nonnegative deconvolution for spike train interference from population calcium imaging. *J. Neurophysiol*, **104**(6): 3691-3704

[10] Lutcke, H, F, Gerhard, F. Zenke, W. Gertner, & F. Helmchen (2013). Interference of neural network spike dynamics and typology frm calcium imaging data. *Front Neural Circuits*, **7.**

[11] Galasso, J. (2023). Squibbing against continuity claims in AI: Why we can't get there from here.

DOI: 10.13140/RG.2.2.33018.06087

[12] ___(2023). Basal Ganglia Grammars. J. Article. *Open Science Framework* DOI: 10.17605/OSF.IO/AQ5GH

[13] __ (2023). ABABABA-Grammars. *Open Science Framework J. Article* 10.17605/OSF.IO/AQ5GH

[14] __ (2019). A Note on Artificial Intelligence and the Recursive Implementation: Note 4 (Chapter in *Recursive Syntax*, LINCOM Studies in theoretical Linguistics, 61.

[15] Marr, D. & T. Poggio (1976). From understanding computation to understanding neural circuity. *Computer science and artificial intelligence lab*. MIT. https://dspace.mit.edu/handle/1721.1/5782

[16] (a).Moro, A., M. Tettamanti, D. Perani, C. Donati, S. F. Cappa, & F. Fazio (2001). Syntax and the Brain: Disentangling Grammar by Selective Anomalies*. NeuroImage* **13**, 110–118. doi:10.1006/nimg.2000.0668.
(b). Tettamanti et al. (2001). Neural correlates for the acquisition of natural language syntax. *Elsevier Science*.

[17] Hansson U., J. Chen, C Honey (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn Sci* **19**(6): 304-313.

[18] Foster, D. (2017). Replay comes of age. *Annu Rev Neurosci*, **40**: 581-602.

[19] Kastanenka, K., R. Moreno-Bote, M. De Pittà, G. Perea, A. Eraso-Pichot, R. Masgrau, K. Poskanzer, E. Galea (2020). A Roadmap to integrate astrocytes into systems neuroscience. *Glia* **68**(1): 5-26.

[20] (a). Nadeau, S. (2021). Basal Ganglia and thalamic contributions to language function: Insights from a parallel distributed processing perspective. *Neuropsych Review***. 31** https://doi.org/10.1007/s11065-020-09466-0
(b). (2012). *The neurological architecture of grammar*. MIT Press.

[21] Kozachkov, L., K. Kastanenka, & D. Krotov (2023). Building transformers from neurons and astrocytes. *Salk Institute for Bio Studies.* https://doi.org/10.1073/pnas.2219150120

[22] Duarte, R., A. Seeholzer, K. Zilles & A. Morrison (2017). Synaptic patterning and the timescales of cortical dynamics. *Curr Opin Neurobiol*, **43**: 156-165.

[23] Khani, A. & G. Rainer (2016). Neural and neurochemical basis of reinforcement-guided decision making. J. *Neurophysiol* **116**(2): 724-741.

[24] Sweatt, J. (2016). Neural plasticity and behavior—sixty years of conceptual advances. *J. Neurochem*, **139:** 179199.

[25] Lau, B., T. Monteiro, & J. Paton (2017). The many worlds hypothesis of dopamine prediction error: implications of a parallel circuit architecture in the basal ganglia. *CurrO Pin Neurobiol*, **46**: 241-247.
[26] Stephan, K., S. Iglesias, J. Heinzle & A. Diaconescu (2015). Translational perspectives for computational neuroimaging. *Neuron*, **87**(4): 716-732.
[27] Moro, A. (2016). *Impossible Languages.* MIT Press.
[28] Marcus, G. (2001). *The Algebraic Mind*. MIT Press.
[29] Marcus, G. & E. Davis (2019). *Rebooting AI*. Pantheon Books
[30] Galasso, J. (2018). Reasons for Movement. 'Four Sentences' (Sentence No. 4). (Chapter taken *from Reflections on Syntax* Peter Lang Publishing. https://www.scienceopen.com/book?vid=556c6e5f-d2cf-431c-9a3f-d3dc6bfa8877

[31] Seyednozadi, Z. et al. (2021). Functional role in N400, P600 in language related ERP studies.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8419728/#:~:text=The%20most%20important%20language%2Drelated%20ERP%20components%20are%20the%20N400,the%20anomalous%20item%20(6).

[32] Renart, A., R. Moreno-Bote, X. Wang & N. Parga (2007). Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Comput*, **19**(1): 1-46.
[33] Mongillo, G., O. Barak, & M. Tsodyks (2008). Synaptic theory of working memory. *Science* **319**(5869). 1543-46.
[34] Panzeri, S. J. Macke, J. Gross & C. Kayser (2015). Neural population coding: combining insights from microscopic and mass signals. *Trends Cogn Sci*, **19**(3): 162-172.
[35] Binder, J. R. Desai, W. Graves, & L. Conant (2009). Where is the semantic system? *Cereb Cortex*, **19**: 2767-96.

[36] DePittà, M. (2019). G-protein-coupled receptor-mediated calcium signaling in astrocytes*. Computational Glioscience*, Springer.
[37] Cunningham, J. & B. Yu. (2014). Dimensionality reduction for large-scale neuronal recordings. *Nat Neurosci* **17(**11): 1500-1509.
[38] Takata, N., T. Mishima, C. Hisatsune, T. Nagai, E. Ebisui, K. Mikoshiba & H. Hirase (2011). Astrocyte calcium signaling transforms cholinergic modulation to cortical plasticity in vivo. *J. Neurosci*, **31**(49): 18155-18165.
[39] Grodzinsky, Y. & A. Santi (2008). The Battle for Broca's Region. *Treds Cogn Sci* **12**(12). 474-480. DOI: 10.1016/j.tics.2008.09.001
[40] (a) Fodor, J. (2001). *The mind doesn't work that way*. MIT Press.
(b) Pinker, S. (1997). *How the mind works*. Norton.
[41] Pearl, J & D. Mackenzie (2018). *The book of why: The new science of cause and effect*. Basic Books.
[42] Hauser, M. N. Chomsky, & T. Fitch (2002). The Faculty of Language: what is it, who has it, and how did it evolve? *Science* **298**. DOI: 10.1126/science.298.5598.1569

[43] Galasso, J. *Recursive Syntax*. LINCOM Studies in Theoretical Linguistics, **61**.
https://www.scienceopen.com/book?vid=9862ac3f-01d2-449c-84f2-e7d8afc08fff
https://www.academia.edu/37650895/Four_Sentences_Opening_remarks_for_Ling_417_Child_Language_Acquisition_Fall_2018
[44] Posner, M. & M. Raichle (1994). *Images of Mind*. Scientific American Library.
[45] Schrimpf, M. et al. The neural architecture of language. Proc Natl. Acad Sci. USA. **118**. https://doi.org/10.1073/pnas.2105646118
[46] Kahneman, D. (2011). *Thinking, Fast and Slow*. Straus and Giroux.
[47] Kumar, S. et al. (2022). Reconstructing the cascade of language processing in the brain using the internal computations of a transformer-based language model. *Bio Rxiv.*
[48] Humphries, M. et al. (2018). Insights into Parkinson's disease from computational model or basal ganglia function. *J. of Neurology, neurosurgery, and psychiatry*, **89.**
[49] Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron*, **88.** 220-234.
[50] Rumelhart, D. et al (PDP processing group) *Parallel Distributed Processing.* MIT Press.

2

Basal Ganglia Grammar*

The Neuronal Substrate as Common-Denominator Interface for Language

Joseph Galasso[1]

So, as the story goes… 'Some forty-thousand years ago (KYA) our perfect anatomy twin (Cro-magnum) got up and *wanted* to speak'. It goes without saying that this is indeed a very recent turn of events in the long trajectory of the development of our species. The notion of 'wanted' is of interest. What is typical of cultural-motivating norms these days is the belief that *capacity* follows *desire*, i.e., you just need to 'want' to do it badly enough for it to 'happen' (as the Nike logo goes, 'Just do it'). Well, in biological evolutionary-developmental (evo-devo) terms, it's quite the reverse: desire follows capacity[2]. The cliché 'You can do anything you put your mind to' means exactly that: first, you must be *able* to put your mind to it. So, some 40-60KYA early modern man *wanted* to speak—just because *he could.* There are several theories and hypotheses about how this came about, with the spectrum ranging from external social interaction (Functionalism) to an internal innately driven language faculty providing a language algorithm (Generative), or, from slow gradual emergence to quick punctuated equilibrium. In any case, there is a *common-denominator interface* which holds for all such theories of language evolution—namely, that there must have been some newly-acquired ability, genetically predisposed or otherwise acquired via maturation (or saltation) which created this unique capacity for language. To be sure, there are plenty of antecedents as found in the evo-devo literature which may serve as evolutionary precursors to full fledge language: e.g., the abilities to mimic, trace, keep tempo, maintain hand-eye coordination, fashion tools, predict from a pattern, follow from analogy, etc. The holy grail of such antecedents is what is often referred to as 'Theory of Mind'—viz., humans' unique ability to empathize, be altruistic, be able to feel how another might feel and think *without recourse to the actual stimuli of the other.* In a sense, this holy grail (which some say leads to human consciousness) amounts to the ability to *categorize & abstract*. There is nothing in our physical nature which would suggest such exquisite behavior—it is an exclusively mental phenomenon.

*****I was lucky enough to come across '*the grammar of the Basal Ganglia*' in reading Lieberman's 2006 model which largely attributes language to human outsized cortical growth with the afforded supplemental space deemed necessary to create overlapping neural circuity connected to the basal ganglia. While Lieberman's work is skeptical of the claim 'language is special' (believing there are anatomically homologous structures found across species), I believe an extension of his 'Basal-Ganglia' model jibs nicely with the Hauser et al. 2002 model which delimits 'broad' forms of the faculty of language (FLb) (voice, semantics, motor-control) with that of 'narrow' forms (FLn)—the latter of which cortical *recursive syntax along with its substrates is the predominant if not the sole component*. (Also see Friederici, Chomsky et al. 2017).

[1] https://csun.academia.edu/josephgalasso (Draft July 17, 2023). I thank Andrea Moro for pc. on this note. [2] See my 'Myth' paper (2020). (There is an interesting caveat to this: Humans' successful desire to fly, over a hundred years ago. In this case, humans were able to circumvent their physiological constraints for flight by creating a 'work-around' away from *bodily* flight towards *mental* flight. Because we could 'think' flight, we did so. In a sense, we highjacked the long-winding path of evolution of 'physicality' with this abstract shortcut of 'mentality': Humans 'thought' to fly,

and so they did (to the heavens and beyond…while always leaving their humanoid forms firmly tethered to the ground). These unique 'workarounds' which fly in the face of evolutionary biology is the greatest feat humans have ever performed—and is tantamount to the recursiveness as found in humans' unique 'tool-making capacity': [a tool [that can make a tool [that can make a tool]]]… *ad infinitum.*

The ideas herein are meant to only serve as a springboard for future researchers whose intention is to largely support a 'basal-ganglia-account' of human's unique ability for recursive syntax.

What follows is a brief research statement of what I think may be a neural 'common-denominator' behind all these inner-workings leading to language—namely, 'the *Basal Ganglia'* (BG), a subcortical region of the brain not dissimilar to what we know and find of the *insular cortex*, which serves as a switchboard-like operator for all inputs coming in from different regions of the brain, to be sorted and allocated—the accumulation of which leads to the manifestation of language (see e.g., Liebermann 2002/2006). Foremost in importance among these BG inner-workings is the ability for *movement.* For example, in Parkinson's disease we see BG neurological impairment leading not only to loss of *physical* 'noisy' movement control (Huntington's dance), but also to deficits in *mental* movement (so-called 'silent' movement) related to (syntactic) movement-based processes such as long-distance dependences exhorted by the brain/mind (i.e., hierarchical long-distant co-indexing which goes well beyond linear-ABABABA grammars, See Galasso 2023a). (Some of the best studies of such loss of mental, silent movement comes from Broca's Aphasia (BA) subjects who show loss of syntactic movement capabilities.[9] Regarding the latter (silent mental movement) there may be similar maturation factors at play in the emergence of early child language syntax)[10] (The Silent-Unspoken 'language of thought' is a perfect example of such silent movement (see Moro's Lecture)).

Suppose that language, for all of its shrouded glory, is nothing more than a species-specific *incidental acquisition*. We can describe Language-Incidental (LI) this way as a set of procedures which got subsequently established via a dual capacity: i.e., the interface between the actual external stimulus (being observed) and its given internalized proprioceptive-rehearsed representation which got strengthen over rehearsal episodes. (It has now become better understood that the *insular cortex* is quite valuable as a rehearsal mechanism and seems to be implicated as a pathway between short-term working memory and long-term memory)[5]. Hence, LI is a holistic 'external-to-internal' mapping which brings together all relevant compartmentalized language-specific tasks found of the brain. The research claim here amounts to little more than the notion that language (LI) is a byproduct (an *exaptation*) of a dual capacity: the capacity of a species to (i) first have the attention to notice some aspect of the environment, and then the ability to (ii) internally map it (via move-based repetitive rehearsal) to some inner-mental processing (a visual screen if you will). In this manner LI resides as a bridge, as a looping interface between this overt-to-covert (back to overt) procedure. (For repetitive rehearsal procedures as found in the *insular cortex,* see Posner et al. p. 124).

---

[9] See Grodzinsky's work regarding syntactic impairment of BA subjects.

[10] See Galasso, From Merge to Move (2016, monograph). The first significant maturational study I ever came across was the Wakefield & Wilcox (1995) study. (Also, see Radford (1990), Galasso (2003) for a maturational theory of child syntax). The idea taken here is that Broca's area suffers delayed onset due to myelination. [5] See my paper 'Squibbing against continuity claims… (2023b) for a discussion on the aspects of *rehearsal* which come to bear on recursive neuronal circuitry. For insular cortex rehearsal, see Posner et al. (p. 124).

For instance, consider language as sound (phonology). Phonology may start out for an individual as a mere prosodic observable piece of sound, which then becomes mapped by the listener (internal) as a procedural/articulation action-plan. For example, consider the word BOY: the fact that a listener notices and perceives the first-order ambient sound gets you only halfway there—in order for LI to get established in this minimal case, say, regarding the initial phoneme /b/, there must also be some second-order internal mapping and silent-repetitive rehearsal in order to secure the implemented language function. There must be silent movement. It seems to be the case that while a young child may initially only focus on the sound of an utterance, what the child is actually gaining access to is a simultaneous, internalized action-plan of both how to pronounce that same sound, and how to go about classifying and storing its content: theoretically, the phonologist amongst us sets out the features, say of the initial consonant 'B' /b/: [+voice], [+bilabial], [+plosive]…but these are merely *feature-tags* set out as labels in order to represent, store and retrieve what had been already rehearsed and subsequently acquired. Other than this tacit knowledge of relevant phonological features involved, the real trick to LI is in the features subsequent rehearsal, storage & retrieval mechanisms. (Kuhl's work on the 'Native Language Magnet Theory 'shows precisely this—where phonemic targets move from broad to tightly narrow 'phoneme-clusters' over an experiential timespan).

This idea has been around for several years: Michel Paradis (McGill University) speculates[11] that the interface implicates the basal ganglia, a subcortical part of the brain:

> *For example, individuals may focus on the acoustic properties of a word while acquiring the proprioceptive programs that allow them to articulate the word, or focus on the meaning of an utterance while internalizing its underlying syntactic form—which is not there to be observed. It is stored implicitly, in that it remains forever opaque to introspection, as is made clear by numerous unsuccessful attempts at characterizing the underlying structure of sentences since the emergence of Chomsky's early work on the representation of syntactic structures. (p 355)*

In other words, the notion here is that while an individual may be focusing on the external stimuli, she is simultaneously mapping the external to the internal (a dual processing). In fact, it may very well be the case that almost all of 'what we "think" we perceive' (for all aspects of the world, language included) is nothing more than our unique ability to move from external observation to internal mapping (*unique in the sense that only homo-sapiens seem to have this capacity to move from first-order environmental stimuli to a second-order abstract procedural categorization). (Nb. In its strong version, this theory would have it that we never 'really' even see the *item (as pure object)* in front of us—we instantly *categorize* it, and proceed to map it onto its internal representational scheme even before there is time enough to appreciate its true substantive quality)[12]. Perhaps, in one extreme sense, we are always *doing language*

[11] Among others. See Paradis (2003) 'Cerebral Mechanisms in Bilinguals' (p 355), in *Mind, Brain, and Language: Multidisciplinary Perspectives* (Eds. Marie Banich, Molly Mack. Lawrence Erlbaum Ass. Publishers).

[12] My favorite example of this can be found in George Santayana's quote: 'A 19th 20th century cognitive perspective would come to sow how cued-representations (Icons) could only represent an individual or item, while a detached symbol could stand for an item without the unnecessarily burdening requirement of external stimulation—the former being triggered by direct, environmental stimuli, the latter by a delayed response of its memory. The sign that once expresses an idea will serve to recall it in the future'. Regarding the vision of an object in front of us, say, an 'orange'. Note how we can't make out the total spheric nature of roundness since the backside of the orange is

(24/7), whether or not we are listening or speaking. The internal language-mapping screen is set to the 'on-button' 24/7—i.e., we are constantly *talking and listening* as a form of *silent movement*[13]. It seems the essential role of the subcortical basal ganglia—a role indeed unique to our newly-emergent species—is to provide a movement-based algorithm allowing this displacement from (i) noticing of item to (ii) mapping, storage & retrieval of category (insofar that this ebb and flow articulates a recursive feedback loop). To my mind, in addition to research presently being carried out on the basal ganglia area (a region which is predominately involved with both physical movement as well as mental movement, language being the quintessential silent movement), *Mirror Neurons* come closest to such an interface, where external stimulation triggers an already mapped trace of the observed behavior (see Corballis 2010, Tettamanti & Moro 2010).

The idea that a singular region of the brain might be largely responsible for language is contested by most neurolinguistics. However, there does seem to be evidence that some form of a 'neuro loop' is required in order for language to take hold, and that this recursive loop is what has evolved in our species, some say as early as 40-60KYA. For this internal inner-play to succeed—'from notice and observation> to mapping> to rehearsal, storage and retrieval…' begs the question of whether there too are neuro-substrates (so-called 'neuro-nets') which undergird such recursive looping. We await future research into recursive neurons and their implications to regions associated with the basal ganglia.

What we currently know of the Basal Ganglia (BG) and its interface implications to language.

1. In defining language as essentially a *recursive* operation, we can tease out operator distinctions between so-called (broad) 'recurrent' ABABABA-grammars of the $(AB)^n$ type (MERGE), [AB], [ABAB], [ABABAB]…, for example what we find in the vertical stacking of lexical/vocabulary learning (noting how linear/adjacency supersedes without hierarchy), versus true language-specific (narrow) 'recursive' embedded hierarchy of the A^nB^n type (MOVE): [A[A[AB]B]B]…[14] of what we find in syntax. (For 'broad vs narrow' Faculty of Language (FL) distinctions, see Chomsky 1995; Hauser, Chomsky & Fitch 2002; Fitch & Hauser 2004. Also see Friederici et al. 2006; Friederici 2008, 2009). We find that when there are deficits in recursive grammar such as embedded clause interpretations, trace-movement operation, etc. (fn3), several studies suggest that the dysfunctions can be regarded as a main etiological result of a malfunctioning BG. What we can speculate of the evo-devo behind the BG human-like operation parallels that of what we speculate of Broca's area—viz., that we believe it is both of a late evolutionary emergence in terms of phylogeny (of our mammalian class), and that it is sensitive to ontological critical periods (regarding its latent child onset). (One such clarion call for most developmental linguists of the maturational persuasion is that 'Broca's area just hasn't come online yet' in the very earliest stages

hidden from view. Only through an internal theory of categorization (= round) does the object come to full view. (See Edwin Abbott's 1884 novella 'Flatland' for such wonders).

[13] It was the linguist Alec Marantz some time ago who put me straight on this idea: his analogy that our language-T.V. antennas are always at working scanning, even when there is no possible language signals to be picked up. Also see Moro's (2020) Lecture on Impossible Languages: (Part 2) 'the strange case of unspoken language'.

[14] See Galasso (2023a) paper on ABABABA grammar. For Merge vs Move, see Chomsky 2005, Galasso 2016.

of two-three-word child syntax; this having to do with latent myelination surrounding Broca neural circuitry).

2. The BG plays a critical role not only in movement (both physical as well as mental), but also in sequential actions (Marsden & Obeso 1994) which must take place (e.g., in sequential movement of motor-control mappings onto speech articulation, or stem+affix decomposition as found in morphosyntax (fix vs speaks) with IPA showing similar final /s/ though with distinctions in stem+affix processing: with final /_s/ in 'fix' being undecomposed/stem, and the final /_s/ in 'speaks' being decomposed e.g., [fIks] vs [[spik]s]…(e.g., Bybee & Slobin 1982; Pinker 1999).
3. The BG also find correlates to the gene known as *FOXP2 (e.g., Vernes et al. 2006, 2007; Shriberg et al. 2006)—this correlate* suggests that demonstrable 'specific language impairments' related to movement are in fact associated with deficits in the neural pathways making up the BG. This is easily extended to work carried out by Grodzinsky and his studies related to Broca's Aphasia (BA) dysfunctions in syntactic movement operations, relative clauses and binding (Grodzinsky 2000; Grodzinsky et Santi 2008)[15].
4. The Dual Routing Model (also called Dual Mechanism Model/DMM) (Pinker & Prince 1988, 1994; Pinker 1999; Ullman 2001, Galasso 2003) shows how only procedural-knowledge systems seemingly tether to BG—with declarative/working memory tethering to medial Temporal-Lobe (TL) brain structures, (such as the hippocampus, and presumably Wernicke's area (W)), while Front Left Hemisphere (FLH) Broca's (B)/Brodmann's areas 44/45 (B44/45) shows the most connection to BG-related movement. ERP studies perhaps best illuminate these corollaries whereby the N400 signals non-BG connections (and are thus tied to semantics and recurrent ABABABA-grammar) while P600 and LAN signatures tie to recursive system FLH/B44,45/BG (see Schlesewsky 2009 for review). Single Mechanism Models/SMM (Rumelhart & McClelland 1986)—of the classic sort celebrated by connectionism and even current Artificial Intelligence (AI) models (See Marcus 2001; Galasso 2019 (Note-4) 2019; Marcus & Davis 2019)—cannot tease out such distinctions as they have no ability to shift from recurrent to recursive operating platforms.[11]
5. According to the above DMM, the processing of regular/recursive morphosyntactic rules (e.g., Number, Tense, Agreement) is supported by FLH/B44,45/BG in contrast with irregular morphology which is sensitive to frequency—with irregular morphologies pertaining to 'ABABABA'-singular-route processing—a form of non-rule/non-symbolic learning associated with recurrent neural networks as found in the TL regions of the brain. This is indeed what we find.
6. On the question of the evolution of BG itself and its unique ability to support language, the consensus in the literature is as follows: (i) That BG situates in a critical subcortical region of the brain which, over its evolution, has become sensitive to inputs/outputs engaged in the diverse spreading of neuronal interactions, (ii) with such complex computations motivating non-linear/sequential time-step releases (of the sort which require hierarchical structured expressions), (iii) and is assumed (I believe rightly) that the increase in (vertical) working-memory space due to this increase in cortical volume spawned these (horizontal) modes of expressions—with flat-language *Recurrent* 'ABABABA-grammars' being attributed to vertical computations

[15] See Galasso (2023a) paper for discussion. [11]
See Galasso 'Squibbing' paper (2023b).

(corresponding the Hauser et al's LFb), while *Recursive '*Basal-Ganglia Grammars' (syntax) being attributed to horizontal computation (corresponding to FLn). Given these assumptions, one can speculate within accepted research norms that not only is 'language special' since the human basal ganglia is special (an LFn version), but that also this insular subcortical mass does share broad evo-devo characteristics with other non-mammalian cortical homologues (an LFb version). There is still no clear evidence that a singular gene (FOXP2) was solely behind this cortical change: it may be the case that cascading events took place, involving a series of other rightly-timed cascading neuro events—e.g., the transcription of multi-genes where protein-modification and neurotransmitter alterations occurred simultaneously, or within a relatively short evolutionary time span).

Conclusion.

The question of whether or not there are selectively dedicated neuro networks specifically designed to promote the capacity of recursive syntax is still an open question. However, some of the recent studies as cited herein suggest that indeed neuronal and substrate structures can and do support move-based operations designed to (potentially) serve recursive hierarchy as uniquely found in human language. We leave it to future evo-devo neurolinguistic research to discover the precise neurological substrates which undergird this BG recursive machine. My best speculation to date is that *pyramidal neurons,* perhaps in conjunction with other peripheral neuronal circuitry, might be the best candidates for recursive grammar, since there is some evidence that such neurons can create the necessary 'looping effects' (e.g., repetitive looping <cortex-BG-thalamus-BG-cortex>) which would be required of such recursive implementation. Extending this proposal even further, it has been suggested that the pyramidal neuron itself is a recursive-generating machine—i.e., that syntax is found in the neuron itself (see David Marr below). This would be similar to the wonderous idea that memory itself finds its way embedded inside the circuitry of the cell (Eric Kandel).

Additional Studies (on-line):

Recursive Neuronal Connections & Reverberating Circuits (David Marr):

http://toritris.weebly.com/ http://toritris.weebly.com/recursive-connections.html

Recursive vs Recurrent neural networks.

https://ai.plainenglish.io/recursive-neural-networks-rvnns-and-recurrent-neural-networks-rnns-2ff6a067ad01

Theory of Cortical Function. (David Heeger)

https://www.pnas.org/doi/epdf/10.1073/pnas.1619788114

References

Banich, M & M, Mack (eds. 2003). *Mind, Brain, and Language*. Lawrence Erlbaum Ass. Publishers.

Boeckx, C. & K. Grohmann (eds. 2013). *The Cambridge Handbook of Biolinguistics*. CUP.

Bybee, J. &. D. Slobin (1982). Rules and schemas in the development and use of the English past tense. *Language*, 58. 265-289.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press.

Corballis, M. (2010). Mirror neurons and the evolution of language. *Brain and Language* 112, 25-35.

Friederci, A. (2009). Pathways to language. *Trends in Cognitive Sciences*, 13. 175-181.

https://www.sciencedirect.com/science/article/abs/pii/S1364661309000278?dgcid=api_sd_search-api-endpoint

Friederici, A. & J. Bahlmann; S. Heim; R. Schubotz; A. Anwander. (2006). The brain differentiates human and non-human grammars. *Proceedings of the National Academy of Sciences* USA, 103 pp 2458-2463.

https://www.researchgate.net/publication/7312241_The_brain_differentiates_human_and_nonhuman_grammars_Functional_localization_and_structural_connectivity/link/541059940cf2f2b29a40f3ba/download

Friederici, A; N. Chomsky; R. Berwick; A. Moro; J. Bolhuis (2017). Language, Mind and Brain. *Nature Human Behavior* 1. 713-722.

Galasso, J. (2003). *The Acquisition of Functional Categories.* IULC Publications, Bloomington, Indiana. https://www.academia.edu/43939141/The_Acquisition_of_Functional_Categories_A_Case_Study

___(2016). From Merge to Move. LINCOM Studies in Theoretical Linguistics, 59. https://lincom-shop.eu/LSTL-59-From-Merge-to-Move/en http://www.csun.edu/~galasso/WhyMove.pdf

___(2019). *Recursive Syntax*. LINCOM Studies in Theoretical Linguistics, 61. (Note-4: A Note on Artificial Intelligence and the Recursive Implementation. 162-210. Note-5: A Note on Broca, FOXP2, and the role of Move: a neurolinguistic Review. 211-215).

https://lincom-shop.eu/LSTL-61-Recursive-Syntax/en

https://www.academia.edu/40929371/Recursive_Syntax_Table_of_Contents_and_Preface

___ (2020). The Myth of Function defines Form. https://ling.auf.net/lingbuzz/005298

___ (2021). *Reflections on Syntax*. Peter Lang Publishing. (Berkeley Insights in Linguistics and Semiotics, Vol. 101). https://www.peterlang.com/document/1062281

___ (2023a). The Recurrent + Recursive Linguistic Mind: ABABABA Grammars and a Note on Child Syntax. https://lingbuzz.net/lingbuzz/007176

___(2023b). Squibbing against Continuity Claims in Artificial Intelligence: Why we can't get there from here: The pursuit of recursive neurons. https://lingbuzz.net/lingbuzz/007354

Grodzinsky, Y. (1986). Language deficits and the theory of syntax. *Brain and Language*, 27 (1).

135-159. https://www.sciencedirect.com/science/article/abs/pii/0093934X8690009X?via%3Dihub ___(1990). *Theoretical perspectives on language deficits*. MIT Press.

___ (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, 23, 1-21.

Grodzinsky, Y. & A. Santi (2008). The battle for Broca's region. *Trends Cogn Sci,* Dec. https://www.sciencedirect.com/science/article/abs/pii/S1364661308002222

Hauser, M & T. Fitch (2003). What are the uniquely human components of the language faculty? In M. Christiansen & S. Kirby (eds). *Language Evolution*. Pp. 158-151 Oxford press.

Hauser, M; N. Chomsky and T. Fitch (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298. 1569-1579. https://www.public.asu.edu/~gelderen/Hauser2002.pdf

Kuhl, P. (see all of Kuhl's work here). Linguistic Experience and the Perceptual Magnet Effect. https://legacy.cs.indiana.edu/~port/teach/641/Kuhl.Iverson.lingc.exptc.percpt.magnt.pdf

Lieberman, P. (2002). On the nature and evolution of the neural bases of human language. *Am. J. of Phy Anthro*. 35. 36-62.

___ (2006). *Towards an Evolutionary Biology of Language*. Harvard University Press.

Also see link below on evolution of language & thought:

(https://www.researchgate.net/publication/297758854_The_evolution_of_language_and_thought

Am J Phys Anthropol. 35: 36-62. https://onlinelibrary.wiley.com/doi/10.1002/ajpa.10171

Marr, D. (2016). Niv Y, Langdon A. Reinforcement learning with Marr. Curr Opin Behav Sci. 2016 Oct;11:67-73. doi: 10.1016/j.cobeha.2016.04.005. PMID: 27408906; PMCID: PMC4939081.

Marsden, C. & J. Obeso (1994). The Functions of the Basal Ganglia and the paradox of stereotaxic surgery in Parkinson's disease. Brain, 117. 877-897.

Marcus, G. (2001). *The Algebraic Mind*. MIT Press.

Marcus, G. & E. Davis (2019). *Rebooting AI*. Pantheon.

Moro, A; & M. Tettamanti et al. (2020) Syntax and the Brain: Disentangling Grammar by Selective Anomalies. Academic Press.

Moro, A. (2016). *Impossible Languages: When Linguistics Meets the Brain.* MIT Press.

(Lecture: 2020) https://euresis.org/wp-content/uploads/2020/11/Moro-ParliamoDiScienza2020.pdf

Paradis, M. (2003). Differential Use of Cerebral Mechanisms in Bilinguals. (In Banich & Mack (eds). op. cit).

Pinker, S. (1999). *Words & Rules*. Basic Books. https://monoskop.org/images/9/9e/Pinker_Steven_Words_and_rules_the_ingredients_of_language_1999.Pdf Pinker, S. & A. Prince (1988). On Language and Connectionism. *Cognition* 28, 73-193.

Posner, M. & M. Raichle (1994). *Images of Mind*. Scientific America Library.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax.* Basil Blackwell, Oxford.

Rumelhart, D. & J. McClelland and the PDP Research Group (1986). *Parallel distributed processing* (Vols 1, 2). MIT Press.

Schlesewsky I, & M. Schlesewsky (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass, 3*(1), 19 58. https://doi.org/10.1111/j.1749-818X.2008.00099.x

Shriberg, L. & K. Ballard; J. Tomblin; et al. (2006). Speech, prosody, and voice characteristics of a mother and daughter with 7, 13 translocation affecting *FOXP2*. *J. of Speech Language, and Hearing Research*. 42 1461-1481.

Tettamanti M; A. Moro, et al. (2005). Basal Ganglia and Language: phonology modulates dopaminergic release. *NeuroReport* 16(4) 397-401.

Tettamanti M, A. Moro (2010). Can Syntax appear in a mirror (system)? *Cortex* 48, 923-935

Ullman, M. (2001). The declarative/procedural model of lexicon and grammar. *J. of Psycholinguistic Research*, 30. 37-69. https://www.polyu.edu.hk/cbs/rclcn/images/cdl_articles/U/Ullman._2001.pdf

Vernes, S. & J. Nicod; F. Elahi et al (2006). Functional genetic analysis of mutations implicated in human speech and language disorder*. Human Molecular Genetics*, 15. 3154-3167.

Vernes, S. & E. Spiteri; J. Nicod et al. (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of *FOXP2*, a gene mutation in speech and language disorders. *American Journal of Human Genetics*, 81. 1232-1250.

Wakefield, J. & M. Wilcox (1995). Brain maturation and language acquisition. *Proceedings of the 19th Annual Boston University Conference on Language Development*, 2 vols. MacLaughlin, J. & M. McEwen (eds), 643-654. Cascadilla Press.