An Uncouth Approach to Language Recursivity

Eleonora Russo & Alessandro Treves

A simple-minded view is presented here on the problem of the origin of language, which dismisses any relation with hitherto unobserved specific language microcircuits in the cortex as well as with gross connectional hierarchies which are seen also in other mammals. In this view, language arises out of a capacity for spontaneous latching dynamics, which emerges when the connectivity of an extensive cortical network, which need not be hierarchical, crosses a critical phase-transition threshold.

Keywords: associative memory; latching dynamics; origins of language; phase transitions; Potts units; storage capacity

1. Cortical vs. Cognitive Organization

Neuroanatomy, proceeding slowly but surely at the tranquil pace of a descriptive science, would have the authority to inform a basic understanding of the neural mechanisms subserving higher cognitive capacities. For the faculty of language, this has not happened. There is a fundamental mismatch between the conceptual structures invoked to describe the complexity of language — parsing trees, hierarchies of grammars, principles and parameters — and those emerging from the observation of the articulation of the human nervous system. Neuroanatomical dynamics unfold over evolutionary time scales of millions of years, and their main organizational principles have been scholarly described for about a hundred years (e.g., Lorente de Nó 1938). Language dynamics, even though in the most stable parametric aspects may stretch over several thousand years (Longobardi & Guardiano 2009), unleash their astonishing power in the rapid acquisition of a language by a child — in a few years.

As a result, linguists tend to ignore taking stock of the stable organization apparent in the human brain, and at times nurture the mistaken expectation that a sudden discovery from the world of biology, like that of the structure of DNA, will at some point revolutionize the relation between language and the brain, and crack the neural codes for syntax. The Language Acquisition Device (LAD), a remarkable abstract construct (Briscoe 2000), might then acquire the semblance of a neuronal apparatus, taken to be hiding, like the Holy Grail, perhaps in one of the frontal sulci, disguised to non-believers as standard cortical circuitry. While the quest for the LAD goes on, allured by reports of quantitative differences and asymmetries in area 44 or 45 (Uylings *et al*. 2006, Amunts *et al*. 2010), it may be

ISSN 1450–3417

useful to briefly review salient features of cortical organization.

1.1. *Origin and Evolution of the Cerebral Neocortex*

Higher mental processes in the human species massively involve the cerebral cortex, though it should be noted, not to the exclusion of other structures such as the cerebellum or the basal ganglia. The cerebral cortex derives from a structure, the pallium or dorsal component of each hemisphere, as it bulges out of the forebrain behind each olfactory bulb, that was presumably common also to the ancestors of reptiles and birds, and whose ancient amniotic phenotype is thought to resemble most closely the cortex of modern reptiles. In the evolution of mammalian lineages, the two dramatic events that separated them from other amniotes were the lamination of the dorsal cortex and the reorganization of the medial cortex into the modern mammalian hippocampus. These two events likely occurred between three and two hundred million years ago, with outcomes that are anatomically clear and functionally obscure, but in any case are very much with us to this day. No further major reorganization has occurred since, common to all mammalian species, and differences in the organization of specific radiations, such as the incomplete granulation of cetacean cortex (Huggenberger 2008), are more in the way of amendments than bright new ideas.

The two remarkably stable traits of a 3-fold differentiated hippocampus and a 3-fold laminated cortex (Fig. 1) seem to us to express the fundamental mammalian *geist*, and the crucial challenge for theories of mammalian neural computation to try and explain; yet we feel rather isolated in our interest for these two phase transitions (Treves 2003, 2004).

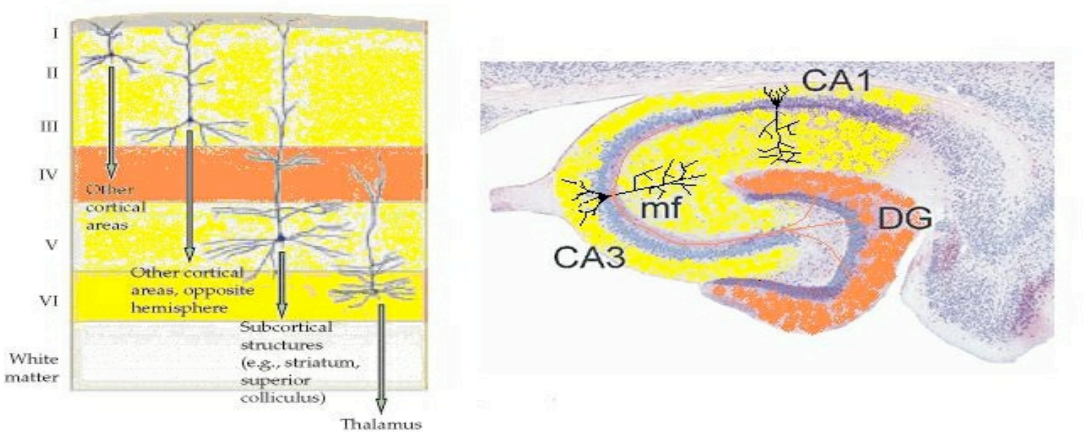
I
II
III
IV
V
VI
White matter
Other cortical areas
Other cortical areas, opposite hemisphere
Subcortical structures (e.g., striatum, superior colliculus)
Thalamus
CA1
mf
DG
CA3

Figure 1: In mammals, the dorsal cortex is laminated (left) and the medial cortex is reorganized into the hippocampus, differentiated into 3 main sub-regions (right). In both structures the essential mammalian innovation is the insertion of an input layer of granule cells (orange) feeding into the pyramidal cells (yellow). Both these traits are common to all mammals, and neither has anything to do with the faculty of language.

Be it as it may, it is hard to identify further qualitative jumps in the structure of our nervous system, beyond these that we underwent together with other mammals. Although the simple-minded notion of cortical uniformity (Rockel *et al.* 1980) has been fiercely criticized (Herculano-Houzel *et al.* 2008, Rakic 2008), modifications and specializations on the basic mammalian design are prevailingly quantitative (Krubitzer 1995, Semendeferi *et al.* 2010). Particularly in relation to the language faculty, they come nowhere close to drawing a boundary, for example, between our brain and that of non-speaking apes, as sharp as can be drawn between, say, bats (mammals) and starlings (birds) — however startlingly recursive starlings may be (Gentner *et al.* 2006).

The dominant neuronal constituents of the cerebral cortex are the pyramidal cells, which are defined by a long axis oriented perpendicular to the cortical sheet. The main new element which sets the mammalian neocortex apart from the reptilian paleocortex is the insertion of a layer of granule cells, layer IV, where many afferent inputs terminate, and which separates infragranular from supragranular pyramidal layers (Fig. 1, left). It is thought that the new arrangement facilitates a precise point-to-point afferent connectivity, enabling the formation, in part by self-organizing processes, of fine topographical maps (Treves 2003). This occurs throughout the expanse of the so-called isocortex, although layer IV is more prominent in sensory cortices, consistent with its role in setting up orderly sensory representations, which become progressively fuzzier in more advanced areas. In frontal cortex, layer IV may remain undifferentiated, suggesting that the cognitive enhancement that accompanied, in the most advanced mammals, the quantitative expansion of the frontal lobes, is not picking up on the technological advance of a laminated cortex. In particular, the faculty of language, which arises recently and only in the human species, appears unrelated to lamination, which emerged hundreds of millions of years ago, in all mammals. It seems doubtful, therefore, to ascribe special significance to limited differences in the exact density of distinct cortical layers (Amunts *et al.* 2004). The concurrent differentiation of the medial cortex into the regions DG, CA3 and CA1 of the hippocampus (Fig. 1, right), with their strikingly idiosyncratic circuitry (Treves *et al.* 2008), is likewise common to all mammals and entirely unrelated to language.

1.2. Cortical Hierarchies

In the reptilian cortex, one distinguishes a lateral, a dorsal and a medial portion, and the olfactory origin of this most anterior component of the brain remains engraved in a clear direction of olfactory information flow, from lateral through dorsal to medial, leading to the old-fashioned but phylogenetically correct interpretation of the hippocampus as the motor division of the olfactory sensory-motor circuit. Even in those mammals, in which lateral olfactory cortex is reduced to little more than a residual, olfactory information from the bulb accesses it directly, a reminder of its former primacy. The other modalities, notably vision, audition and somatic sensation, which long ago penetrated and colonized the cortex from their more posterior original stations, send afferent inputs not directly but through the thalamus to the dorsal cortex, and there join the flow towards the medial component. In more complex mammals these new 'settlers' are seen to have established an increasing number of visual, auditory and somatosensory maps, leading to the great expansion of the dorsal cortex. Such maps, including advanced processing areas in which topography has become so vague as to have been erstwhile overlooked (Kaas 1997), are laid out in a hierarchy from peripheral (primary sensory) towards central/medial. The hierarchy finds unambiguous expression in the arrangement of the cortico-cortical connections: Lower (more peripheral) areas send inputs mainly from their supragranular layers towards the layer IV (whether or not populated by granular cells) of higher areas, which send, mainly from their infragranular layers, back projections to lower areas, which terminate there in layer I. This connectivity pattern allows neuroanatomists to determine which of two areas is higher up in the hierarchy.

Originally the pallium, like the dorsal aspect of the spinal chord, is sensory. The motor cortex is thought to have differentiated relatively late, with mammals, from parts of the somatosensory cortex. Does the motor cortex, which has no layer IV, sit on top of the cortical hierarchy defined by cortico-cortical connectivity? Hardly so. It is the hippocampus, and other limbic components adjacent to it, that the laminar origin and termination of the projections elect as their supreme leaders (Barbas 1986). In fact, both in the temporal and in the frontal lobes it is the more lateral areas, which are more laminated, that project axons that terminate deep into the medial temporal or caudal orbitofrontal cortices respectively, which therefore are higher order in the connectional sense (Rempel-Clower & Barbas 2000). Information flows towards the limbic system, towards establishing memories, whereas the stream leading to motor cortex and the expression of overt behavior is more of a diverted back projection towards the periphery. It should be noted that these connectivity patterns are not rigorous rules but rather statistical trends, and that they may just reflect fossil functionalities, simply because no more modern organizational principle has come to supersede them. Even in this perspective, that is a connectional hierarchy we certainly have.

What are the changes that have then occurred in our neocortex, over this last couple of hundred million years? The most striking one, across several mammalian species, is in size, accompanied however by the parcellation of the cortical expanse into an increasing number of areas, in what is known as the process of arealization. Areas are defined by sometimes very subtle differences in cellular composition or laminar organization, that make the cortex resemble a patchwork of tonalities of the same hue, more than a blanket with continuously shaded colors. In sensory regions it appears very clearly that boundaries between areas are determined by the edges of topographic sensory representations, whether or not salient histological differences are seen with the neighboring area. In so-called simple mammals, a sensory modality may be represented by one or two topographic maps, before information is fully mixed with that from other modalities; in complex and more arealized mammals like ourselves, tens of distinct areas, each a complete map of sensory space, may represent a single modality. Several areas present further granular or quasi-granular structures, such as the mini-column, hyper-columns and pinwheels of cat and monkey visual cortex or the barrels of rat somatosensory cortex (Kaas 1997). These traits, sometimes popularized as evidence for a generic columnar organization principle, appear instead to be specializations that, in certain areas and in certain species, refine *ad* *hoc* the common design expressed by laminated neocortex (Rakic 2008); the proliferation of distinct areas, in contrast, appears as a universal option available across species and modalities, and utilized more by some and less by others.

What keeps cortical areas together? Cortico-cortical connections. Unlike local connections, comprised by axons that never leave the gray matter or by the early departing collateral branches of those that do, cortico-cortical connections travel through the white matter, linking together areas that may sit at different levels of the periphero-hippocampal hierarchy, or at about the same level. Since cortico-cortical connections linking areas at about the same level, as well as local connections, are not hierarchical, and local and non-local connections are estimated to come in roughly equivalent numbers, it is likely that strictly hierarchical connections are a minority, unlike what is assumed in certain neural network models. The cortex is largely democratic, and a given unit can usually find itself, depending on the circumstances of life, pre-synaptic or postsynaptic to another unit, or both. Causal reasoning, which informs many conceptual models of cognitive processes, is manifestly inadequate to capture the web of potentially reciprocal influences that cortical neurons (and cortical areas) exert on one another.

Whereas a democratic arrangement of all cortical neurons on the same footing is unique, many distinct ways can be conceived of ordering them in a strict or loose hierarchical arrangement (Fuster 2009). There are, in fact, at least two more ways to define a connectional hierarchy, beyond that based on the laminar pattern of connections between areas. One is to focus on the number of synaptic contacts on the basal dendrites of pyramidal cells, that Guy Elston (2000) has estimated to increase dramatically going from occipital to temporal and then frontal cortex. Basal dendrites receive mainly local recurrent excitation, so the observation suggests a posterior-frontal gradient from more input-driven to more recurrent circuits. Another way is to focus on the density of terminals of neuromodulators, in particular dopamine, which is particularly high in prefrontal cortex. This indicates a shift from a more rigid, operationally stable processing mode in posterior cortex to something more subject to multiple modulating influences in the front. Note that a dopamine gradient is seen also in birds, in the near absence of a full-fledged (and in any case mono-layer) cortex. These connectional/anatomical hierarchies therefore partially overlap with each other, but only in a loose sense, and there is no evidence that they are evolutionarily related, or geared towards a common functional purpose.

1.3. *Task Switches and Prefrontal Cortex*

The scientific study of animal behavior has been in part based on the experimental paradigm of classical conditioning, whereby, since Pavlovian times, an animal is exposed to stimulus x followed, in rapid temporal succession, by another stimulus y. In its 'operant' variant the animal learns instead that to stimulus x it must respond y, to get reward z (or to avoid a punishment). In their crude simplicity such paradigms lend themselves to quantitative measures, more than ecological behavior, and have been extended in several directions. A trivial extension is to increase the number of stimulus-response associations, for example, to x_1 the subject should respond y_1 to get z, whereas to x_2 the subject should respond y_2, to x_3, y_3 and so on. Another type of extension, also increasingly used with human subjects, involves the combinatorial articulation of contingencies, for example, in context w_1, to x_1 one should respond y_1, and to x_2, y_2, whereas in context w_2, to x_1 one should respond y_2, and to x_2, y_1. The paradigm can obviously be complexified by expanding the number of associations, but also by adding levels, for example, in task v_1 when in context w_1, to x_1 respond y_1, while in w_2, to x_1 respond y_2, and vice versa in task v_2. Obviously, if elements v, w, x are purely labels, their complete configuration can be recapitulated in a compositional variable, say u, where, for example, u_1 denotes 'molecular' configuration (v_1, w_1, x_1), u_2 denotes configuration (v_1, w_2, x_1), etc. If, however, elements v, w, x are presumed to have a life of their own, both in the real world and as represented in the brain of subjects, it is convenient to maintain an atomic notation, to point out that a certain stimulus–response association, for example, $x_1 - y_1$, is correct in task v_1 and in context w_1, but not in situation (v_2, w_1). The experimenter can add levels of contingency n, n+1, n+2, *ad libitum*, making the paradigm progressively more complex.

A recurring observation from the analysis of brain lesioned patients and from neuroimaging studies is that the most anterior portions of the cortex appear to be involved in the correct learning and execution of the higher contingency levels, with perhaps the frontal pole necessary for the maximum complexity level $n_{\max}$, in a particular type of paradigm, that normal human subjects are able to deal with. While there is no evidence that $n_{\max}$ is universal across different paradigms, the review of neuroimaging data and the *ad hoc* experiments designed by Badre & D'Esposito (2007) indicate that $n_{\max}$ can be higher than 3, in the sense that they can distinguish at least 4 hierarchical levels of contingency processing in a series of paradigms of increasing complexity.

One may represent such paradigms as a tree, where each variable is a branch generating at the end node the various thinner branches it can lead to. For example, branch w_1 'generates' x_1, x_2 e x_3, which in turn 'generate' y_5 and y_8. The tree representation misses out the combinatorial nature of the process, because if a branch represents x_1 generated by w_1, a distinct branch shall represent x_1 generated by w_2, and yet another x_1 generated by v_1: The tree structure does not allow for multiple parents. Branches must then multiply, and more branches be assigned to the same event when produced by distinct 'causes'. The apparent complication is counter-balanced by the logical clarity of the tree structure, which allows analyzing contingencies as in a chess game. Also in chess, the same move may follow distinct moves by the other player, or one's own, but a mental tree representation may facilitate an assessment of the current situation, at the price of some redundancy. When mentally climbing on a branch that corresponds to exactly the same situation of the pieces as another already visited branch, we only need to identify the two and retrieve the configuration of thinner branches, and leaves, already explored.

Inconsistencies may arise if, following Badre & D'Esposito, to the branches of a tree representation, conceived as descriptive of a mental process, one wishes to associate neuronal activity in certain cortical areas so that, for example, value x_1 of variable x implies specific activity by a particular group of neurons. And one insists on a generically valid correspondence, so that those neurons 'code' for x_1.

Then either the distinct branches corresponding to x_1 when generated by distinct causes are represented by the activity of distinct groups of neurons, and then the tree hierarchy is clear but the coding is confusing and highly redundant, or they are all represented by the same group, in which case the coding is clear but the hierarchy loses much of its general significance, beyond the individual experimental design, and the simple-minded logical tree of contingencies grows into a mangrove of multi-factorial events, eventually sublimating into a web of interactions of surreal complexity — a cortical network, obviously.

1.4. Syntactic Trees and Hierarchical Processing in Language

A domain in which tree representations have been developed to more powerful sophistication and have offered what seems like an essential contribution is, of course, in the description of syntax in natural languages. Such phenomenologically observed syntax is often interpreted, in the various streams of formal linguistics, as the imperfect biological manifestation of an exact underlying structure, which takes the form of an (upside-down) parsing tree. The terminal branches, or leaves, are associated roughly to individual words *w, x, y, z* of a natural language, while the non-terminal branches are associated to grammatical constructs *A, B, C* that do not appear overtly in natural language, but which are usually construed to have a neuronal representation of some form in our brain. From a start symbol *S* one generates a sentence by following not a *single* branch at each node, as in the complexified conditioning paradigms mentioned above, but *multiple* branches, as specified by certain rewrite or production rules. The corpus of sentences, generated by all possible ways of following each production rule, coincides conceptually with the language and can be represented by a gigantic tree, but even a single sentence corresponds to a tree that has as many leaves as, roughly, words — in contrast to single chess games and individual conditioning trials, which do not correspond to flourishing trees, but rather to destitute trees that have lost all leaves but one. There is thus a hierarchy of levels of analysis implicit in each sentence, if described by a parsing tree, which does not necessarily match the hierarchies necessary to parse other sentences, even within the same formal framework.

Chomsky (1955) has famously shown that such frameworks can be classified in a further, abstract hierarchy of frameworks, from unrestricted to context-sensitive to context-free to regular grammars. This was a beautiful achievement, fertile with insightful connections to then-developing computer science. Strictly speaking, it is more of an ordered set of inclusion relations than a *bona fide* hierarchy. (It does not befit a group of oligarchs to be much more numerous than the populace they rule over — if it is unrestricted grammars that are considered to be on top; and it does not belong to the rulers to be simpler-minded and less powerful than their subjects — if vice versa it is regular grammars that are considered to be on top.) In any case, the hierarchy of grammars has not been associated to hierarchical structures in the brain, except perhaps by some fundamentalists who have reputed simpler brains (which they might unwittingly ascribe to non-human primates, or to other mammals) to be capable of cognitive processes equivalent to regular grammars.

Parsing hierarchies, instead, have been associated, in one of the most striking results of applying linguistic theories, with an orderly progression in the severity of the deficits in aphasic patients with different patterns of agrammatism (Friedmann 2001). In particular, among three groups of agrammatic patients, the one most impaired patient could be described as being able to access only the leaves and the thinnest branches of the syntactic tree expressing the structure of a Complementizer Phrase (CP), which, in the usual upside-down scheme, has the CP node at the top and the leaves dropping down at various levels on the left. 13 severely impaired individuals could be described as being able to climb up two more levels, and 5 mildly impaired one to climb two further levels up, getting almost within sight of the CP node, as it were. These intriguing observations suggest the psychological reality of parsing trees, as ways to order syntactic structures in terms of relative complexity; they do not however, point to a correspondence between those trees and connectional hierarchies in the cortex, even though one is tempted to make that inference, given the long standing neuropsychological association between specific brain lesions and specific behavioral impairments.

The relation between agrammatic impairment and brain lesion is made problematic by the possibility, in general, that the same production rule be represented, in different derivations (in the parsing of different sentences) at different levels of the tree. Then, if different levels of the parsing tree are forced to correspond to different levels in one of the connectional hierarchies of the cortex, either one assumes that a given production rule be potentially expressed by the oper-ation of several different cortical areas, or else one has to assume that the anatomo-functional correspondence is itself variable, from sentence to sentence, with maybe only the initial symbol S, which generates the whole sentence, represented in a stable manner in a mythical spring, somewhere in the frontal lobes, out of which every sentence gushes forth. Neither option is particularly appealing, once made explicit, but in implicit form they may guide our thinking, however unwilling we are to acknowledge it.

2. Recursion and Recurrence

Recursion, in whatever form, makes any attempt to impose a rigid hierarchical processing scheme on the cortex even more difficult, in the same plain sense in which social mobility disrupts rigid social hierarchies. Recursion in a weak form arises immediately when one conceives of a natural language as being satisfactorily approximated by the corpus generated by a finite set of production rules, in the sense that the same set of rules is applied at each step k in the generation of a sentence, however long, and potentially the same terminals (the leaves) can be attached at different steps. This might be dismissed as trivial recursion, but modelers who take seriously the challenge of identifying neuronal mechanisms apt to implement syntactic operators, for example, in terms of fillers and roles (beimGraben *et al*. 2008; Borensztajn *et al*. 2009; Battaglia & Pennartz, in press; see also Namikawa & Hashimoto 2004), devote much of their creativity to dealing with such 'trivia', and appropriately so. This form of recursion can stretch out across very many steps, especially when syntactic dependences are considered to extend, as they do in real life, beyond individual sentences. It is, to all intents and purposes, infinite recursion.

Recursion takes a stronger form when the set of rules allows for choosing the very same individual rule at different steps, as necessary to model simple complementizer sentences like 'John reports that Mary says that…' or, as in Dante's 13th canto of the Inferno, 'Cred' ïo ch'ei credette ch'io credesse che tante voci uscisser, tra quei bronchi…'.

Yet more complicated forms of recursion occur when instead of a linear chain, that, for example, includes two non-terminal elements *A* e *B* (Verb Phrases and Noun Phrases, say), which take terminal values $a_1b_1a_2b_2$ a_3b_3, one has an embedded structure like $a_1a_2a_3b_1b_2$ b_3 or even $a_1a_2a_3b_3b_2$ b_1. Formal grammars that admit these structures are the less restricted ones, but it seems obvious that such structures are needed in order to model natural language without having to resort to byzantine constructs. The design of artificial systems that produce language-like strings endowed with such embedded structures is difficult, and it has been suggested that non-human primates cannot speak because they cannot manage the embedded type of recursion (Fitch & Hauser 2004). Yet there is no convincing evidence that non-human primates can acquire, and speak fluently, natural languages even without embedded structures (Hauser *et al*. 2002). And from the point of view of messing up the correspondence between processing along a fixed anatomical hierarchy and along a dynamically rearranged syntactic tree, embeddings are not needed: The simpler types of non-embedding recursion are sufficient — especially when recursion is expressed along several distinct dimensions. A point about embeddings which is often overlooked is that humans generally have trouble understanding, and rarely produce, embedded structures with more than 3 or 4 levels of embedding. Thus the distinction between infinite recursion in humans and finite in other species, proposed by Hauser *et al*. (2002) does not coincide with that informing the experiments of Fitch & Hauser (2004), between finite levels of embedding in humans and zero in other species.

What the observation of even simple forms of recursion suggests is that one should abandon the hypothesis that, during language production, processing should occur along an ordered hierarchy of cortical areas, whether or not specialized for distinct operations (as 'modules' in the Fodorian sense; Fodor, 1983). Such hypothesis originates in computer science and in the block diagrams of early cognitive psychology, but is completely foreign to the world of cortical information processing, where recurrence is the rule. If humans were to process information along a feed-forward series of stations when they speak, they would be singularly handicapped, given that all other mammals, and other amniotes, and humans when they do not speak, use complex recurrent circuits all the time. It would be very odd if recurrent processing were to be silenced only in order to permit, of all things, a prominently recursive functionality such as language!

It seems, therefore, that excess reliance on artificial intelligence approaches to language, on trying to analyze language as it would be if it had evolved among computers and not among humans, has led astray the search for the neuronal mechanisms of language production, which conceivably have to be found among generic cortical mechanisms. But perhaps with a twist.

3. Quantity May Produce Quality

What distinguishes the human cortex from that of other mammals are its dimensions. In the past, it was thought that the human brain was larger than that of other mammals only in relation to body weight, because in absolute terms, elephants (5 kg) and whales (8 kg) have larger brains than humans (1.4 kg). Recently, however, Suzana Herculano-Houzel (2009) has discussed the possibility that the human cortex may have more neurons, also in absolute number, than any other mammal (and any other living organism). Her argument is based on the observation that the human cortex appears to scale up linearly with respect to other primates, with an approximately constant density of cells per unit volume; whereas with rodents the scaling is strongly sub-linear: A large rodent like the Capybara has much reduced density, and many fewer cells than expected in a linearly scaled up mouse of that body weight. While the scaling laws for the density of neurons in proboscidea and cetaceans are not known, it is likely, she argues, that they will end up making the largest whale and elephant brains less dense, as with rodents, resulting in total number of neurons in their cerebral cortices around 3 billion, compared to 16 billion in the human cortex. So, the human cortex would be the one with more neurons, after all.

Other quantitative parameters that affect the capabilities of neuronal networks are those that determine their connectivity. The observation of uniform design principles for the cerebral neocortex, across mammalian species and cortical areas, should not be taken to imply that the 'canonical' cortical circuit (Douglas & Martin 1991) is exactly the same, hence tacitly assume it to operate always in the same manner (Rakic 2008). Guy Elston, in fact, has argued now for a number of years that important quantitative differences exist in the number of spines present on the dendrites of pyramidal cells. Focusing on estimating the number of spines on basal dendrites, taken to be indicative of the number of independent recurrent synaptic inputs from other pyramidal cells nearby, he has reported much lower numbers in occipital than in temporal or frontal cortex, e.g., 1,000 vs. 7,000 or 9,000, respectively, in the macaque. Further, the corresponding numbers are all significantly higher in the human cortex, ca. 2,000, 13,000, and 15,000, respectively (Elston *et al*. 2001). Such quantitative differences are normally overlooked in conceptual reasoning, but they can easily produce qualitative differences in the functionality of a neural network.

3.1. Phase Transitions

The physics of phase transitions offers a poignant model for how quantitative differences in some parameter describing a complex system of interacting units can generate major qualitative differences in the collective behavior of the system. This will not be reviewed here, but it suffices to note that conceptual causal reasoning alone, where consequences are logically associated to qualities, with disregard for quantities, would have had difficulties explaining why water is liquid at 33° F, but turns into ice just 2 degrees below, or why a ferromagnetic material can suddenly lose its properties upon heating. Conceptual explanations, sadly still the dominant epistemological paradigm in the cognitive sciences, are inadequate when dealing with phase transitions. The mathematical techniques originally developed to analyze models intended to describe one particularly complex type of materials exhibiting phase transitions, the so-called spin glasses, have instead been successfully adapted to analyze models of associative memory networks, following the suggestion by John Hopfield (1982). Such networks, like spin glasses, exhibit phase transitions, for example, when their effective 'temperature' — a measure of the variability ascribed to noise — changes a bit, but also with changes in their connectivity (Amit *et al*. 1987, Treves & Rolls 1991). Attractor states representing memory items then disappear, if either the effective temperature is too high or the connectivity too low in relation to memory load, and the network enters a phase in which it cannot function as an associative memory. This is the phase transition associated with storage capacity, which we can denote with the critical memory load — the maximum number of memory items p_c above which associative retrieval fails.

Most studies of formal/mathematical models of associative memory have focused on single networks, which have often been interpreted as representing a patch of cortex, for example, 1 mm^2, containing of order 10^5 neurons. Valentino Braitenberg, however, has proposed considering the entire cortex, or at least its fronto-temporal associative areas, as an "associative memory machine" (Braitenberg & Schüz 1991), including, in the human brain, of order 10^5 patches with 10^5 pyramidal cells each (Braitenberg 1978; see Fig. 2). One can consider a mathematical model of such a two-tier associative memory network, in which neurons are grouped into compartments with dense internal connectivity and sparse connectivity between compartments. A model of this type can be called modular, but not in a Fodorian sense, since all the modules, representing real patches but as if they had sharp boundaries, have the same structure and mode of operation. It shows phase transitions, at least the storage capacity phase transition, in that the asymptotic attractor states correlated with each of p stored memory patterns are only present when both the internal connectivity and the one between modules are sufficiently dense, in relation to p (O'Kane & Treves 1992).

Specifically, the memory patterns can be defined across the entire network as composed of local patterns stored in each module, which can store S local attractors if the number of internal connections per unit, C_s, is sufficient. The number of long range connections per unit, C_l, has to be sufficient to support the retrieval of the p stored combination of local attractors, against the interference of all other possible combinations. The minimal values for C_s and C_l that allow for successful retrieval, given p and S, depend in a complex manner on the parameters characterizing the architecture of the network and the memory representations it encodes (Fulvi Mari & Treves 1998). This makes it convenient to analyze a simplified model, in which the lower tier of the modules with their internal connectivity is replaced by a symbolic representation, in terms of variables which can take multiple values standing for the multiple local attractors of the full model (Fig. 2).

These variables are called Potts units, and can be thought of as tiny vectors pointing towards the multiple directions of a hyper-pyramid in S dimensions, with its vertex at the origin, a vertex that stands for the inactive state of the local module. Such a Potts version of an associative memory network, which then explicitly models only the upper tier of the two-tier architecture, has first been analyzed by Kanter (1988) and then by Bollé *et al*. (1991, 1993). They considered discrete Potts units, which can be in a state or another but not partly in a state and partly in another. Graded-response units have been considered later (Treves 2005) and they allow a more realistic modeling of local patch dynamics, including firing rate adaptation, i.e. neural fatigue, a pervasive feature of cortical dynamics.

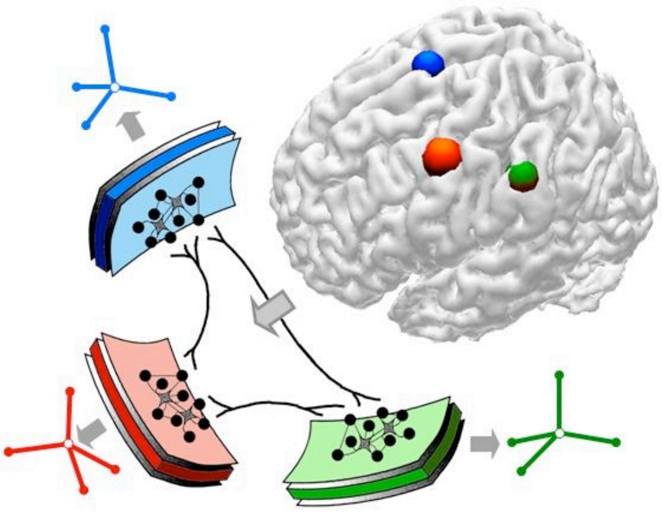

Figure 2: From the cortex to Braitenberg's 'skeleton' model (larger arrow) to the Potts model (small arrows), through abstraction and simplification. Each Potts unit has S 'attractor' states (filled) plus the 'quiescent' state (empty circle).

3.2. *Latching Dynamics*

Including a model of firing rate adaptation leads, in the presence of *correlations* among the global attractors, to a new phenomenon: latching dynamics. Latching dynamics is the hopping of the network from one attractor state to another, where the first, while decaying away due to neuronal fatigue, acts as a cue to retrieve the second, due to their being correlated. The process can be repeated a few times or even indefinitely, in which case one talks of infinite latching.

Latching dynamics are not recursive *per se*. If the transitions from one attractor $\sigma(n)$ to the next are random, there are no rules of the type $\sigma(n+1) = \Omega[\sigma(n)]$ being recursively applied at each transition. If the transition probabilities are non-trivially *structured*, however, either sculpted by a learning process or at least embedded in the correlations between attractors, then the transition matrix $\Omega[\sigma]$ can be regarded as implicitly recursive. It was in fact shown that even for a simple non-structured Potts network the transition probabilities are structured, by the correlations, and 3 distinct classes of transitions can occur between attractor states (Russo *et al*. 2008). More interestingly, it was shown that these transitions cannot be described by a first-order Markov process, as they depend on preceding states, much as words do in natural languages (Russo *et al*. 2011), so one has to think in terms of long probabilistic dependencies σ(n+1) = Ω[σ(n), σ(n–1), σ(n–2), …]. The Potts network has memory, in its spontaneous gibberish.

3.3. *The Phase Transition to Infinite Latching*

We have analyzed latching dynamics in a number of studies (Kropff & Treves 2006, Russo *et al*. 2008, Russo *et al*. 2011) to which we refer for a more detailed description of this behavior of the Potts associative network model, and of its possible relation to neurophysiological observations. The point we want to note here is that latching may never occur, terminate by itself after a few steps, or continue indefinitely. Once the parameters of the Potts network are set, the exact duration of the process, and whether it terminates or not, depend on the exact initial conditions. In a large network, however, the dependence on the initial conditions may become negligible, and if latching terminates after a while it has a well-defined length, dependent only on structural parameters. One can then talk about a phase of finite latching (which may include a region of zero duration, where latching does not even start) and a phase of infinite latching.

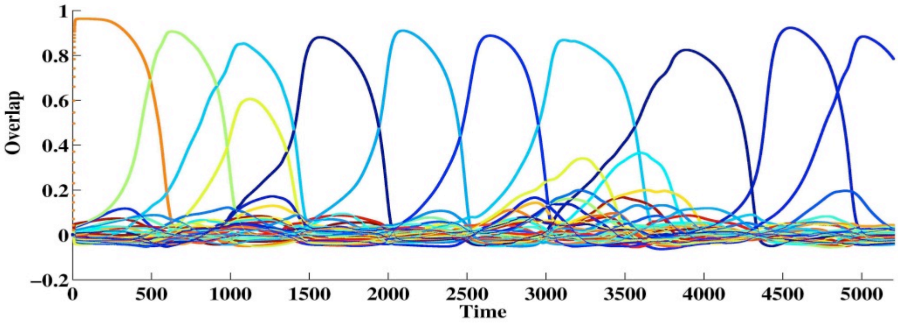

Figure 3: Latching dynamics are shown by plotting the time course of the overlaps of the state of the Potts network with each of p memory states (in different colors). Ideally, the network hops from attractor to attractor, with rapid transitions and protracted permanence in each attractor. In practice, with many choices of parameters, the observed dynamics are much noisier, a rumble-and-tumble with occasionally several overlaps simultaneously high, and many others significantly above chance level. Here, an input cue is presented at time 0, and then the network is left to its own dynamics. Time steps are in arbitrary units, but interpreting them as msecs, or as fractions of msecs, may be useful to suggest a correspondence with cognitive time scales.

While our detailed analysis of the boundary between the phases of finite and infinite latching will be discussed elsewhere, it is useful to sketch the scenario that emerges from the possibility of an abrupt phase transition — a distinct phase transition from the storage capacity one, which we had labeled with the critical storage load p_c. The network in the two phases is identical, and the only difference is in the numerical values of some parameter, for example the number S of local states of each patch/Potts unit or the connectivity C of the Potts units (related to the long-range connectivity in the underlying two-tier model, C_l). A small change in the value of the parameter then opens the gate for a distinct emergent property, which is manifest in one phase and not in the other. In the Potts model of the two-tier cortical network there are therefore two critical boundaries. First, there is a boundary between retrieval and non-retrieval phases. The network cannot function as an associative memory because it cannot retrieve a memory item with a partial cue, when the storage load p is above a value p_c proportional to C and to S^2 (Kropff & Treves 2005). Second, there is a boundary between finite and infinite latching. The network latches indefinitely when the memory load is *above* a certain critical value p_l, because correlations among attractors increase with the memory load. At least in a certain parameter regime, simulations indicate that p_l does not depend on C and scales up with S, approximately linearly. A phase space for the network has to be drawn in (at least) 3 dimensions, p, C and S, and two orthogonal sections though this phase space are shown in Fig. 4.

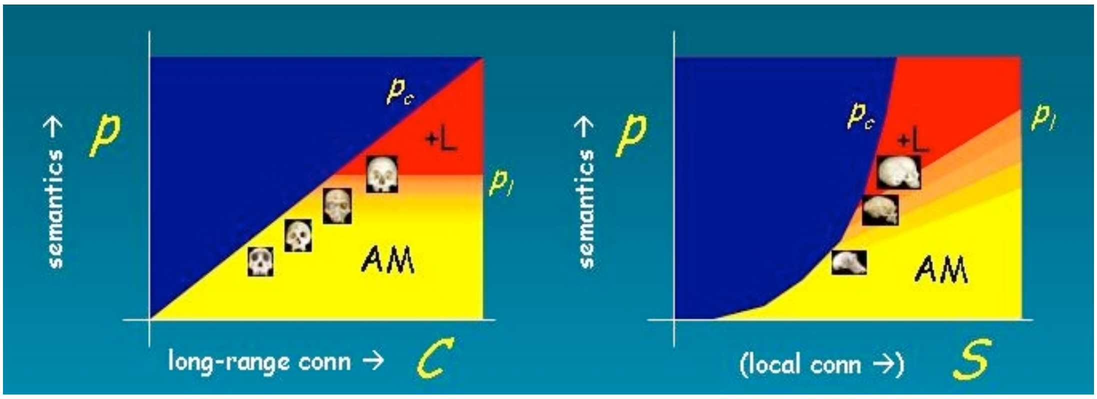

Figure 4: Two schematic cross sections through the 3D phase space of a Potts associative memory. In the p-C section, left, the network can operate as an associative memory (AM) below a critical value p_c which is approximately linear in C. Above a value p_l, which does not seem to depend on C, the network also latches indefinitely (+L). A quantitative increase in long-range connectivity, at some point in the evolution of the human brain, may have triggered the emergence of indefinite latching. A similar phase transition (also depicted as a graded change through orange colors, because its exact nature has not been completely determined yet) from the AM to the +L phases may be seen in the p-S section, right. In this case p_c is taken to be quadratic in S, and p_l linear.

Our on-going analyses indicate that the phase diagram of even simple unstructured Potts networks is not as simple as suggested by Fig. 4. Still, the scenario remains open, that a slowly evolving quantitative increase in the connectivity of the cortex may have suddenly crossed a critical threshold, in the human species, several tens of thousands years ago, that brought the cortical network into a phase characterized by long spontaneous latching sequences, or by their real cortical equivalent, without altering the intrinsic make-up of the network or any of its constituent properties. Latching is an emergent property, or a somewhat more complex set of emergent properties, which emerge when crossing certain thresholds.

4. Conclusion

It is a long shot to extrapolate from transitions produced by randomly generated correlations among attractors, in a crude Potts network, to the recursive concatenation of linguistic structures in speech, even though latching transitions have been shown to display a certain degree of internal complexity (Russo *et al.* 2008). Still, the Potts model indicates the possibility that a recursive mechanism may emerge through a phase transition, in a manner entirely unrelated to the hypothetical appearance, in evolution, of a novel piece of neural circuitry with specific language-adaptable properties (the mythical LAD), or to the refinement of specific connectional hierarchies among cortical areas. The latter may of course encroach on the originally non-hierarchical mechanism and complexify it, but they (the LAD and the hierarchy) may have nothing to do with the emergence of the mechanism. Further studies are needed in order to understand how latching dynamics can be sculpted in a more purposeful manner than by randomly generated correlations, through for example temporally asymmetric synaptic plasticity.

References

Amit, Daniel J., Hanoch Gutfreund & Haim Sompolinsky. 1987. Statistical mechanics of neural networks near saturation. *Annals of Physics* 173, 30–77.

Amunts, Katrin, Marianne Lenzen, Angela D. Friederici, Axel Schleicher, Patricia Morosan, Nicola Palomero-Gallagher & Karl Zilles. 2010. Broca's region: Novel organizational principles and multiple receptor mapping. *PLoSBiol* 8, e1000489.

Amunts, Katrin, Axel Schleicher & Karl Zilles. 2004. Outstanding language competence and cytoarchitecture in Broca's speech region. *Brain and Language* 89, 346–353.

Badre, David & Mark D'Esposito. 2007. Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience* 19, 2082–2099.

Barbas, Helen. 1986. Pattern in the laminar origin of corticocortical connections. *The Journal of Comparative Neurology* 252, 415–422.

Battaglia, Francesco P. & Cyriel M. Pennartz. In press. A computational architecture for memory consolidation: A correlational code for episodic memory training hierarchical semantic networks. *Frontiers in Computational Neuroscience*.

beimGraben, Peter, Dimitris Pinotsis, Douglas Saddy & Roland Potthast. 2008. Language processing with dynamic fields. *Cognitive Neurodynamics* 2, 79–88.

Bollé, Désiré, Roland Cools, Patrick Dupont & J. Huyghebaert. 1993. Mean-field theory for the Q-state Potts-glass neural network with biased patterns. *Journal of Physics* A 26, 549–562.

Bollé, Désiré, Patrick Dupont & Jort van Mourik. 1991. Stability properties of Potts neural networks with biased patterns and low loading. *Journal of Physics* A 24, 1065–1081.

Borensztajn, Gideon, Willem Zuidema & Rens Bod. 2009. The hierarchical prediction network: Towards a neural theory of grammar acquisition. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 2974–2979.

Braitenberg, Valentino. 1978. Cortical architectonics: General and areal. In Mary A.B. Brazier & Hellmuth Petsche (eds.), *Architectonics of the Cerebral Cortex*, 443–465. New York: Raven.

Braitenberg, Valentino & Almut Schüz. 1991. *Anatomy of the Cortex: Statistics and Geometry*. Berlin: Springer-Verlag.

Briscoe, Ted. 2000. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language* 76, 245–296.

Chomsky, Noam. 1955. The logical structure of linguistic theory. Ms., Harvard University/Massachusetts Institute of Technology. [Published in part as *The Logical Structure of Linguistic Theory*, New York: Plenum, 1975.]

Douglas, Rodney J. & Kevan A.C. Martin. 1991. A functional microcircuit for cat visual cortex. *Journal of Physiology* 440, 735–769.

Elston, Guy N. 2000. Pyramidal cells of the frontal lobe: All the more spinous to think with. *Journal of Neuroscience* 20, RC95 (1–4).

Elston, Guy N., Ruth Benavides-Piccione & Javier DeFelipe. 2001. The pyramidal cell in cognition: A comparative study in human and monkey. *Journal of Neuroscience* 21, RC163 (1–5).

Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

Friedmann, Naama. 2001. Agrammatism and the psychological reality of the syntactic tree. *Journal of Psycholinguistic Research* 30, 71–90.

Fulvi Mari, Carlo & Alessandro Treves. 1998. Modeling neocortical areas with a modular neural network. *Biosystems* 48, 47–55.

Fuster, Joaquin M. 2009. Cortex and memory: Emergence of a new paradigm. *Journal of Cognitive Neuroscience* 21, 2047–2072.

Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.

Hauser Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.

Herculano-Houzel, Suzana. 2009. The human brain in numbers: A linearly scaled-up primate brain. *Frontiers in Human Neuroscience* 3, art. 31, 1–11.

Herculano-Houzel, Suzana, Christine E. Collins, Peiyan Wang, Jon H. Kaas & Roberto Lent. 2008. The basic nonuniformity of the cerebral cortex. *Proceedings of the National Academy of Sciences USA* 105, 12593–12598.

Hopfield, John J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences USA* 79, 2554–2558.

Huggenberger, Stefan. 2008. The size and complexity of dolphin brains — a paradox? *Journal of the Marine Biological Association of the UK* 88, 1103–1108.

Kaas, Jon H. 1997. Topographic maps are fundamental to sensory processing. *Brain Research Bulletin* 44, 107–112.

Kanter, Ido. 1988. Potts-glass models of neural networks. *Physical Review* A 37, 2739–2742.

Kropff, Emilio & Alessandro Treves. 2005. The storage capacity of Potts models for semantic memory retrieval. *Journal of Statistical Mechanics: Theory and Experiment*, P08010.

Kropff, Emilio & Alessandro Treves. 2006. The complexity of latching transitions in large scale cortical networks. *Natural Computing* 6, 169–85.

Krubitzer, Lea. 1995. The organization of neocortex in mammals: Are species differences really so different? *Trends in Neuroscience* 18, 408–417.

Longobardi, Giuseppe & Cristina Guardiano. 2009. Evidence for syntax as a signal of historical relatedness. *Lingua* 119, 1679–1706.

Lorente de Nó, Rafael. 1938. Architectonics and structure of the cerebral cortex. In John F. Fulton (ed.), *Physiology of the Nervous System*, 291–330. New York: Oxford University Press.

Namikawa, Jun & Takashi Hashimoto. 2004. Dynamics and computation in functional shifts. *Nonlinearity* 17, 1317–1336.

O'Kane, Dominic & Alessandro Treves. 1992. Short- and long-range connections in autoassociative memory. *Journal of Physics A: Mathematical General* A 25, 5055–5069.

Rempel-Clower, Nancy L. & Helen Barbas. 2000. The laminar pattern of connections between prefrontal and anterior temporal cortices in the Rhesus monkey is related to cortical structure and function. *Cerebral Cortex* 10, 851–865.

Rakic, Pasko. 2008. Confusing cortical columns. *Proceedings of the National Academy of Sciences USA* 105, 12099–12100.

Rockel, A.J., Robert W. Hiorns & Thomas P.S. Powell. 1980. The basic uniformity in structure of the neocortex. *Brain* 103, 221–244.

Russo, Eleonora, Vijay M.K. Namboodiri, Alessandro Treves & Emilio Kropff. 2008. Free association transitions in models of cortical latching dynamics. *New Journal of Physics* 10, art. 015008.

Russo, Eleonora, Sahar Pirmoradian & Alessandro Treves. 2011. Associative latching dynamics vs. syntax. In Rubin Wang & Fanji Gu (eds.), *Advances in Cognitive Neurodynamics II*, 111–115. New York: Springer.

Semendeferi, Katerina, Kate Teffer, Dan P. Buxhoeveden, Min S. Park, Sebastian Bludau, Katrin Amunts, Katie Travis & Joseph Buckwalter. 2010. Spatial organization of neurons in the frontal pole sets humans apart from great apes. *Cerebral Cortex*, doi: 10.1093/cercor/bhq191.

Treves, Alessandro. 2003. Computational constraints that may have favoured the lamination of sensory cortex. *Journal of Computational Neuroscience* 14, 271–282.

Treves, Alessandro. 2004. Computational constraints between retrieving the past and predicting the future, and the CA3-CA1 differentiation. *Hippocampus* 14, 539–556.

Treves, Alessandro. 2005. Frontal latching networks: A possible neural basis for infinite recursion. *Cognitive Neuropsychology* 21, 276–291.

Treves, Alessandro & Edmund T. Rolls. 1991. What determines the capacity of autoassociative memories in the brain? *Network* 2, 371–397.

Treves, Alessandro, Ayumu Tashiro, Menno P. Witter & Edvard I. Moser. 2008. What is the mammalian dentate gyrus good for? *Neuroscience* 154, 1155–1172.

Uylings, Harry B.M., Annelise M. Jacobsen, Karl Zilles & Katrin Amunts. 2006. Left-right asymmetry in volume and number of neurons in adult Broca's area. *Cortex* 42, 652–658.

Eleonora Russo
SISSA
Sector of Cognitive Neuroscience
Via Bonomea 265
34316 Trieste
Italy
russo@sissa.it

Alessandro Treves
SISSA
Sector of Cognitive Neuroscience
Via Bonomea 265
34316 Trieste
Italy
ale@sissa.it