Title
Steps to a semiotic cognitive neuroscience

Author
Deacon, Terrence W

Publication Date
2022-09-23

DOI
10.4324/9781003051817-4

Peer reviewed

Steps to a semiotic cognitive neuroscience

Terrence W. Deacon, Department of Anthropology, University of California, Berkeley, CA

email: deacon@berkeley.edu

ORCID ID http://orcid.org/0000-0002-8672-5461

Abstract:

Finding common ground between neuroscience and semiotics requires reframing neurological processes as semiotic processes and vice versa. In the neurosciences this requires abandoning both the computational metaphor and the view that the basic functional unit of cognition is an individual neuron's response properties. In semiotic theory this requires abandoning a structuralist conception of the sign relation and adopting a process perspective that treats iconic, indexical, and symbolic relations as phases in a hierarchically recursive interpretive process. To achieve this I argue that the basic neurosemiotic operation is iconic interpretation and that neurosemiotic interpretants take the form of neural dynamical attractors.

Key words: representation, affordance, interpretant, counter-current, recursion, icon

I. Cognition explained away

Semiosis: the missing link

Breakthroughs in neuroscience have accelerated at an incredible rate. This has been aided by new tools for observing brain activity *in vivo* during all manner of cognitive tasks. Using animal models to analyze neural function at the cellular and molecular level has provided incredible detail concerning the function of neurons and the connections between them. Cognitive science has also expanded to become one of the fastest growing areas of experimental research. It has revolutionized what once was a largely qualitative science by taking advantage of the logic of computer algorithms to precisely analyze, simulate, and test models of mental processes. So, both at the molecular-cellular and functional levels cognitive neuroscience seems poised to uncover some of the deep mysteries of brain function. And yet, this accumulation of ever-increasing detail about brain functions and cognitive processes still seems no closer to bridging a critical explanatory gap: the link between neural activity and mental experience.

In many respects the promise of cognitive science has been predicated on the assumption that the computational metaphor could either explain or explain away this experiential feature of cognition. And yet there has always been a dilemma lurking at the very heart of this paradigm. Computation is ultimately just a re-description of a machine's operation, in which each machine state is assigned a representational interpretation and each distinct semiotic operation is mapped to a machine operation. From a computational perspective each representational relation is simply assigned by fiat. It is the product of a separate extrinsic re-description assigned by a designer/user, and not anything intrinsic to the machine. So, applying this metaphor to brain function passes the explanatory buck to an already assumed interpretive perspective. Following this to its logical implications can only result in an eliminative conclusion (i.e. that representation is epiphenomenal) or a pan-psychic/pan-semiotic conclusion (i.e. that all physical processes are also semiotic processes). Both effectively collapse the mind/matter distinction and successfully pass the explanatory buck.

Cognitive neuroscientists widely use the term “representation” in the literature, but without a clear definition (Vilarroya, 2017). Worse, by setting aside the messy issues involved in understanding the nature of representational relationships, cognitive explanations tend to alternate between and combine two alternative computational paradigms, often characterized as algorithmic (symbolic) processes and neural network (subsymbolic) processes. This distinction, based on a generic conception of symbolic representation, has a long prehistory that is entangled with the history of logic, mathematics, and linguistics. One aim of the present analysis is to re-complexify this conception of ‘symbol’ and thus expose the hidden infrastructure that computational approaches ignore.

Nevertheless, treating cognition algorithmically and in terms of connectional networks has had a pragmatically valuable effect: cognitive neuroscience has been remarkably productive precisely because it didn’t need to confront underlying semiotic issues. So long as the problem of explaining representation could be provisionally set aside and treated as an unanalyzed mapping relation, the correlations between brain processes and cognitive activities could be studied with the same sorts of analytical tools as used by other physical sciences. Within this framework, identifying a correlation between engaging in a particular cognitive task and observing heightened neural activity in an anatomically distinct brain region has become the gold standard of cognitive neuroscience. How the neural activity is causally related to the phenomenal experience of that process is left for philosophers to ponder.

But this amounts to a methodological dualism. Is this unavoidable? I don’t think so. It’s just that studying the relationship between neural processes and mental experiences has been impeded by lack of a bridging language in which both physical and semiotic attributes can be co-defined. In this respect, semiotic theories might have the potential to help overcome this methodological dualism, since it is the goal of semiotic analysis to characterize different modes of representational relationship and explain the processes that give rise to them. In other words, semiotic theories directly address the very issues that cognitive neuroscience has bracketed. Unfortunately, semiotic theory has mostly been explored within the humanities; debated in philosophy, and employed in the arts. What might be described as “semiotic science” is for the most part undeveloped. So, this potential has, until now, mostly gone unrealized.

If we have any hope of reversing this trend and to develop tools that are useful for analyzing neuro-cognitive functions in semiotic terms we must on the one hand deconstruct some of the non-semiotic assumptions that have become the unspoken foundation of contemporary cognitive neuroscience and on the other hand find ways to ground semiotic theory in physical processes.

The non-semiotic framing of cognitive science

In a computational framework any inferential process that can be completely described in terms of the manipulation of sign-tokens can be instantiated by the operations of a machine.

This is often described as the Church-Turing conjecture (for review see Copeland 2020) and became the default assumption of computational cognitive science after Newell and Simon (Newell and Simon 1972, 1976). The idea was that machine operations could substitute for the corresponding human mental operations to manipulate representational tokens or their equivalents. Presumably, all that would required would be an appropriate mapping between meaningful sign-token manipulations and transformations between machine states that preserves sufficient isomorphism.

An advantage of the computational paradigm, that enabled it to eclipse the behavioristic paradigm that preceded it during the first half of the 20th Century, was that algorithms were structures interpolated between inputs and outputs that could be of indefinite complexity. Unlike introspective descriptions or behavioral input-output mappings, algorithms can be precisely described and analyzed to any arbitrary level of detail. If any act of cognition can be modeled as the manipulation of representational states according to some algorithm then it would seem that cognition can be studied empirically by just analyzing the syntax of the algorithm.

But like behaviorism, this algorithmic conception of cognition effectively bypasses any need to explain the experience of mental representation ("aboutness"), and therefore is agnostic about the semiotic aspects of cognition. In many respects the computational model of cognition can be conceived as an extension of behavioristic logic into the hidden structures that mediate between inputs and outputs. This is nicely exemplified by the recently popular "neural net" approaches to cognition.

Consider, for example, a typical neural net classifier that can be trained to recognize spoken English words as distinct from nonword sounds. The system is “trained” by providing the analogues of rewards and punishments in the form of strengthening and weakening connection strengths with respect to their correlated involvement in contributing to correct or incorrect classifications. A fully trained neural net is effectively an algorithm that is structured to produce the desired input-output mapping given complex input data. Representation is reduced to a mapping relation that is implicitly prefigured in the pattern of extrinsically assigned correct and incorrect correspondence relations.

A recently attractive alternative to computationalism is provided by dynamical systems approaches to cognition (e.g. Ward 2002). But they too are largely antagonistic to any theory of mental representation. They instead invoke the “structural coupling” of brain dynamics with environmental dynamics. Although few would disagree with the view that the immediate neurological correlate of sensory experience is a change in the dynamical activity of certain brain structures, the fact of some dynamical “coupling” alone has little explanatory value. The challenge is to determine what specific complex sort of dynamical relationship constitutes the neurological competence to recognize and act with respect to the only partially predictable constraints and dynamics of the environment. Dynamical coupling does not distinguish simple efficient causal relationships from representation relationships, and yet it is precisely the *non efficient aspects* of semiotic relationships that demands explanation. So, like the computational metaphor, this too evades rather than addresses the problem of explaining mental representation.

The remainder of this essay sketches the outlines of a theory that reinterprets semiotic theory in neurophysiological terms and reanalyses brain functions in semiotic terms.

Part II begins by shifting the focus of semiotic analysis to the process of generating interpretants—mental sign production—and away from treating sign vehicle properties as the basis for their referential function. It focuses on iconic interpretation as the basic neurosemiotic operation from which all higher order semiotic relations are derived.

Part III takes up the challenge of mapping this semiotic interpretive logic to the neurological processes that support them. It argues that iconic interpretant generation involves evoking dynamical attractor patterns of neural activity within distinct cortical areas, and explores the relationship between cortical architecture, interpretant differentiation, and the restructuring of mnemonic processes by symbolic language.

II. Semiotic preliminaries

Semiosis: the process of interpreting

In many respects semiotic theory is the complement to computational theory. Its primary concern is to explain the nature of representational relationships and to make sense of the different modes of representation and how they are interrelated. Central to this approach has been the reframing of representation in terms of the concept of 'sign.' This shift in emphasis was introduced by the father of modern semiotic theory, Charles Sanders Peirce (1839 - 1914), in part to avoid the psychologism of the late 19th century in which the concept of mind was taken as an unanalyzed container of ideas and assigner of meaning. Sign relations for Peirce are discrete phenomena that can be characterized by distinctive physical and relational properties that determine their relationships to one another and how this constitutes their referential and pragmatic effects. But what is most relevant for cognitive science is the way that this shift of emphasis breaks down the distinction between communication and thought. A central claim of Peircean semiotics is that "We think only in signs." (CP 2.302[1]) This implies that there should not be a different set of concepts used for studying minds and studying communication.

From this perspective cognition *is* semiosis: the process of producing and interpreting signs. In this way Peirce described the process of mental interpretation in terms that didn't invoke the actions of a mind but rather only as a process involving the production of signs. In these terms interpretation is a physical process with a specific form. Signs are not reducible to mere tokens, as in computational paradigms, and interpretation is not reducible to the mere manipulation of tokens. The physical tokens are not signs in themselves, but are sign vehicles (the term I will use here instead of Peirce's neologism '*representamena*'). A sign is for Peirce a complex relation between three components: a sign vehicle, that to which it refers (which he called its *object*), and another sign that establishes this link (which he called an *interpretant*; about which more will be said below). But the physicality of all of these phenomena, their relationships to one another, and their production in the interpretive process suggests that a

[1] Standard volume and paragraph reference for the Collected Papers of C. S. Peirce (1931-58).

bridge between the physical and representational aspects of mental processes might be possible in these terms.

According to this perspective, the process of interpretation is the second-by-second development and differentiation of sign relations that emerges with the generation of signs in response to other signs. It is a process that can be observed both outside and inside of interpreters. In this respect, the signs that are labeled in our various taxonomies are not abstract categorical types. From a cognitive point of view they should rather be understood as transient phases in a process of interpretation. This suggests a departure from the way semiotic terminology is most commonly used. In the analysis of semiotic relationships in the arts or humanities, for example, it is common to label certain events or tokens as signs of a particular type. But when the focus turns to the interpretive process, as is necessary to adapt semiotic analysis to mental processes, this structuralist tendency is unhelpful. For semiotic theory to be useful for cognitive neuroscience, semiotic classification must ultimately be regulated by a theory of the generation of semiotic relationships in which the homunculus of an interpreter is replaced by a process analysis. The necessity of this approach is implicit in Peirce's introduction of the concept of an interpretant.

An interpretant is a sign generated in the process of interpreting another sign. It is a mediating sign that is necessary to endow a given sign vehicle with its link to something it can thereby represent. Thought is in this sense a physical sign generation process. It is a process that may be entirely confined within the brain, be coupled to muscular activity, or involve the interactions of agents. The central role played by interpretant production makes semiotic relationships retrospective in an important sense. Although the presence of a sign vehicle precedes the generation of its interpretant in time, its semiotic function (i.e. that which makes it a sign rather than a mere event or artifact) is assigned post hoc by this process of interpretant generation. In this respect the semiotic function of a given sign vehicle only comes into existence after the process of interpretant generation is completed. Although superficially it might appear that the interpretant necessarily follows the sign in time, this way of phrasing things makes the mistake of assuming that a given sign vehicle is already a sign *of something* before it is interpreted. This does not imply, as is sometimes suggested, that because an interpretant is also a sign vehicle that interpretation is indefinitely postponed. The interpretant is not the interpretation. Rather, it is the physical process of interpretant generation itself that transforms a non efficient possible relationship into an efficient actual relationship. The semiotic process is a normal physical process but one that introduces new efficient causal relationships into the world.

Semiotic affordances

Iconic reference depends on shared formal properties between the sign vehicle and what it refers to. For example, with a bit of imagination a face can be discerned on the full moon, or in a cloud formation, and this experience might even remind you of someone you know. But between any two phenomena there are often innumerable similarities. The fact of shared form therefore cannot determine iconism. It is the interpretive use of a particular similarity that matters. Moreover, simply failing to notice a distinction can be sufficient to ground iconic reference, as when a color blind person confuses unripe and ripe fruit. This shows us that it is not the similarity in the sign vehicle and object that matters, but rather whether the interpretations given to them differ. The determinative role of interpretive competence is particularly evident in the case of highly complex and abstract forms of iconism. Consider for example a mathematical equation. Peirce argues that an equation refers iconically, but only once you know how to discern its symbol-mediated isometry (e.g. between the structure of the equation and a corresponding geometric or dynamical relationship). It can only be interpreted to be iconic (e.g. of a parabolic trajectory), if one knows how to discern the way that progressive differences in the results of operations on the various symbolic elements of the equation correspond to progressive differences in the positions of points constituting the geometric object of reference. Thus, recognizing its holistic iconicity is the conclusion of a multistep interpretive process that may extend over a considerable period of time. It is necessarily preceded by a number of iconic, indexical, and symbolic interpretive stages, in the process of constructing this final complex interpretant.

Indices refer by correlation in space, time, or or contiguity deriving from a common substrate. A simple correlation can therefore be the ground for indexical reference. So, for example a lipstick smear on a man's shirt collar can be a troublesome indication to his wife, a urine scent on a branch can be a sexual index to a female lemur, and the mobbing call of a small bird can indicate the present of a raptor. What gets correlated and how (physical, cultural, or evolutionary, respectively) is arbitrary, but the *fact of correlation* is not. So a rat in a Skinner box that has been trained to press a bar in response to a bell in order to get a water reward has learned that the bell is an index of the state of the apparatus even though this pairing was the whim of an experimenter. The arbitrary pairing doesn't make the one a symbol of the other.

For the rat what matters is discovering this correlation so that it can use the bell to indicate the availability of water. Though the particular physical properties of the bell sound have no intrinsic relationship to water, learning the reliable temporal correlation with the water reward makes it an index. Analogously, repeated pairing over the course of generations in the evolution of a species can similarly be the basis for an innate tendency to interpret something indexically. This is the case for the indexicality of a vervet monkey alarm call sound, its correlation with the presence of a type of predator, and the urge to engage in an appropriate defense activity or escape behavior (like climbing tree to avoid a leopard).

What matters, then, is the interpretive process and how it makes use of selected sign vehicle attributes. In the terms of ecological psychology (e.g. Gibson 1966) the attributes of sign vehicles that are selected to bring to mind a given object of reference are merely "affordances." Depending on the interpretive process, they can be used or ignored.

Deconstructing the code analogy

Since the beginnings of philosophical thought human cognition has been analyzed in terms of our most familiar publicly available aid to thought: language. Not only did linguistics contribute critically to the initial development of cognitive science, along with computational theory, many cognitive theories have been explicitly framed in linguistic terms. This is only natural since language is the principle means by which we convey our thoughts to one another.

But treating language as a generic mode of cognition inverts the developmental and evolutionary relationships between language and other communicative and mental functions. In evolutionary terms language is an unprecedented, most recent, most divergent, most highly differentiated form of animal communication and its role in organizing cognition is likely also a recent overlay on much more ancient and basic mental processes. In semiotic terms as well, language is dependent on a complex infrastructure of more basic semiotic processes.

Ironically, the study of language and its neurological underpinnings is an area of cognitive science research that has long suffered from a lack of semiotic sophistication. A major reason for this is that the unprecedented symbolic mode of referring is often treated as though it is self-evident and simple, when it is in fact the most distinctive and atypical mode of reference to have evolved in the animal kingdom. Failure to recognize that symbolization is not just generic representation collapses a critical distinction that obscures its dependence on a vast infra symbolic foundation.

In most linguistic texts symbolization tends to be treated as an arbitrary (unconstrained) mapping relation between two classes of objects—signifiers (e.g. words) and what they signify (e.g. concepts). From this perspective, linguistic reference appears as simple as possible, involving minimal defining criteria compared to other forms of reference. When conceived in these terms, symbolic reference tends to be defined negatively; as reference that does not depend on either iconic or indexical criteria; thus arbitrary. From this perspective symbolic reference seems primary and both iconic and indexical reference forms are treated as the more complex and derived forms.

But this superficial characterization is deeply flawed. This negative way of defining symbols only tells us that neither formal likeness nor factual correlation are used as the basis for symbolic reference. This fails to specify exactly how the symbolic referential relation is established. It merely passes the buck, so to speak, to some assumed and unexplained means by which reference is established, often described as convention (discussed in more detail below).

First consider the nature of "arbitrarity." All semiotic relationships include some degree of arbitrarity, because those attributes that are taken as the ground for the sign-object linkage can be chosen from many dimensions, and only some will be utilized. In this respect, anything can be chosen as a sign vehicle to represent anything else by any mode, depending on the interpretive process involved. For example, depending on the interpreter and social context the same facial grimace can be interpreted as a sign of pain, a neurological disorder, or the communication of social censure. Or the sight of smoke can be seen to resemble a cloud, to indicate a fire, or to signal the election of a pope. The relevant property is selected by the interpreter and what is being referred to depends on the capacity of the interpreter to use that property in a particular way.

So although it has become common to call conventional sign vehicles *symbols* this shorthand terminology has obscured a critical distinction. The current vernacular that calls alphanumeric characters *symbols* only reflects the fact that these characters have been explicitly developed over millennia to express symbolic relations. When your computer begins randomly spewing alphanumeric characters onto your screen they are interpreted as indications of a malfunction, not as symbolizing anything.

Equating symbolic reference with mere conventional correspondence also ignores a complex semiotic infrastructure, since convention itself is a consequence—not a cause—of semiotic processes. Conventions don't just exist by fiat. They are shared interpretive habits that must be created or discovered. Social conventions in particular need to be created semiotically, i.e. by communicating. To avoid circularity, then, we need to understand how conventional forms of communication are created using nonconventional semiotic processes.

As Charles Peirce pointed out over a century ago, we must distinguish properties of a sign vehicle (e.g., a drawing or spoken word), from properties taken to link it to its object of reference. Symbolic reference involves a conventional type of sign vehicle that additionally represents its object of reference in a conventionally-mediated way. Something can be considered symbolic, then, only if the property determining its relationship to what it refers to is also due to a shared rule or convention. Symbols are in this sense doubly conventional. They are conventionally derived sign vehicles that refer to things via conventionally determined means. But this means that they must be derived from other sorts of referential relationships that are involved in the creation of these two levels of conventionality, and these must ultimately rest on a nonconventional foundation.

So instead of using the term 'symbol' to refer to a conventional sign vehicle, as is common in a computational paradigm, semiotic theory reserves the term 'symbol' for sign forms that are doubly conventional. This helps distinguish symbolic reference from conventional iconic (e.g. the typographical smiley face) and conventional indexical reference (e.g. a military salute), as well as from nonconventional iconic (a face in the clouds) or indexical (a loud bang) relations. And because of this doubly conventional aspect, symbolic reference is necessarily more complex and more demanding of interpretive work than iconic and indexical relations. To construct these conventional frameworks presupposes antecedent semiotic work in which prior iconic and indexical processes are required. This means that to explain the neurological basis of symbolic interpretive competence it is necessary to have already explained how brains interpret things iconically and indexically.

Nested interpretive dependencies

Probably the most important contribution of semiotic theory to the analysis of mental functions will derive from recognizing the hierarchically nested dependency of the three canonical modes of representation relationships: icon, index, and symbol. They are most often treated as mutually distinct sign categories defined by the relationship between the properties of the sign vehicle and the properties of its objects (described above). However, this taxonomic classification ignores a deeper constructive relationship between these modes of reference. This is because these defining affordances aren't what determines the mode of reference. As noted above, anything can be taken as a sign representing anything else in any respect (e.g. either icon, index or symbol) if an appropriate interpretation process is generated. Neither the intended interpretation of a produced sign nor any intrinsic properties of a given sign vehicle determines its semiotic function. Intrinsic sign vehicle features can at best constrain and bias interpretation. It’s the form of the interpretive process that matters. But also as noted above, the interpretation process consists in the production of other signs—interpretants—and it is the pattern of this interpretive sign production process that determines how a given sign relation is differentiated.

For an example of this hierarchic interpretive process consider the process of interpreting the significance of a windsock (see figure 1). Imagine, if you will, a naïve traveler observing a windsock through a window, having never seen one before. What interpretive steps would be necessary for her to recognize that it indicates the strength and direction of the wind? Her interpretive competence will depend on knowing that fabric and similar materials (e.g. clothing, umbrellas, palm trees, sails, etc.) will tend to hang vertically unless something lifts them. It will also depend on the experience of seeing flags (etc.) being blown horizontal by a strong wind. Observing the windsock stretched out will thus tend to bring the memory of these experiences to mind, if only dimly and vaguely. These remembered experiences are initial interpretant signs invoked by their similarity in form to the form of the extended windsock. Although invoked spontaneously without effort, their generation is the recognition of an iconism (maybe only experienced[17] as a sense of vague familiarity). But each of these memories also has something else in common with the others: an association with the experience of wind. This is an additional secondary iconism shared between these initial interpretants. It might be described as an *insight* because it is effectively a perception of features of the perceptual process itself. So this additional iconism among the invoked memories comprises a higher order iconism that is the interpretant which must be generated in order to infer that the windsock indicates something not currently experienced: the strength and direction of wind.

Without recognizing an initial iconism between the windsock and these remembered experiences, and without recognizing the additional iconism also shared by these remembered experiences—the presence of wind in each instance—there could be no such inference and no indexical interpretation of this invisible inferred correlate of the windsock's behavior.

[PLACE FIGURE 1 HERE]

But this interpretive capacity can also be made more complex. Discovery of the indexicality of a windsock can be augmented with additional iconic and indexical features. The distinctive tubular form of the cloth-like material and its placement on a pole so that it can freely rotate can be given an additional interpretation. Whereas the initial recognition of this indexical relation is merely a unique instance, these other features of the windsock can be seen as non-incidental to its intended semiotic function. The windsock might therefore also be recognized as a conventional device designed specifically for this indicative purpose. Its distinctive form and its presence in a context where wind strength and direction might have special relevance (e.g. an airport) can now further contribute to interpreting the windsock as a conventional type of sign rather than as merely an accidental one-of-a-kind occasion that provides this information.

Of course, all of this mental activity necessary to achieve this indexical inference can take place in a fraction of a second. The memories don't need to be rendered in any detail, just barely differentiated to the point that their higher order iconism is thereby evoked. And with experience, as this interpretive habit is developed, the entire process will be distilled by a kind of least-work tendency to strip away all unnecessary detail so that this interpretive infrastructure becomes automatic and nearly unconscious. And yet, without recognizing an initial iconism between features of the windsock and these remembered experiences, and without recognizing the additional iconism between these remembered experiences (i.e. that wind was also part of each experience), there can be no such implicit inference and no indexical interpretation.

Notice that in this example, interpreting the windsock as an index was a function of generating multiple levels of iconic interpretants; icons in response to the windsock and higher order icons in response to certain similarity relationships among these icons. The indexical interpretation is in this sense a sign relation of a higher order than the iconic interpretations upon which it is based. This logic can be generalized as follows: *interpretation involves the production of lower-order signs that bring a more developed sign into existence, with iconic signs representing the lowest order.* Thus, every sign is the result of a developmental progression through a series of recursive phases of increasing semiotic complexity until the process stops because of its sufficiency for a given pragmatic end. Both where the process starts and where it stops determine a level of iconism; the point at which further distinctions can be ignored.

To generalize the point of this example, the transition from one interpretive phase to the next higher phase is made possible by the generation of lower-order signs (mentally generated interpretants) along with their relationships to one another. This additional multiple sign generation process required to interpret a newly experienced sign comprises a complex interpretant (which may involve many steps) that mediates between the experienced sign and what it refers to. Thus an internal mental sign production process (interpretant generation) can create increasingly more developed signs.

III. Semiotic neuroscience

A dynamical basis for cortical iconism

Understanding the hierarchic logic of interpretant generation has two important neurosemiotic implications: first, it suggests that the fundamental neurosemiotic process is iconic interpretation;[2] second, it suggests that the generation of higher-order forms of sign relations (like indexicality) is the result of recursively generating iconic interpretations of other iconic interpretations. In neurological terms, this suggests that the basic neurological operation is the generation and recognition of iconic relationships and that there needn't be neural structures specialized for indexical and symbolic functions. These higher order semiotic processes are instead a function of differences in the interrelationships between different modalities and levels of iconic interpretation processes. It also suggests that the process of generating iconic interpretants is at the same time both a perceptual and behavioral activity, because generating a pattern of neural activity can both interpret sensory input and be linked to extrinsic effectors. For this reason it shouldn't surprise us that the basic cellular architecture of the cerebral cortex doesn't radically differ from sensory-specialized to motor-specialized cortical areas,

[2] The idea that the generic function of cerebral cortical circuits is iconic is similar to the theory that cortical circuits have evolved for memory storage, recall, and recognition that is presented in the 2005 book *On Intelligence* by Jeff Hawkins (the inventor of the PalmPilot) with co-author Sarah Blakeslee.

except in the relative numbers of cell types in the different layers and patterns of afferent innervation and efferent targets. It also suggests that in many respects perception is a form of action and action is a form of perception.

To claim that the fundamental semiotic function of a cortical area is iconic is only partly helpful. Even if we accept this hypothesis, it begs what is still one of the most fundamental questions of neuroscience: How is this iconism realized neurologically?

In the 1960s, David Hubel and Torsten Wiesel used single cell recording to demonstrate that individual neurons in the cat visual cortex were differentially responsive to distinct features in distinct positions within the visual field. By1981, when Hubel and Wiesel were awarded the Nobel Prize in Medicine for this work, single neuron recording was considered the gold standard for understanding neural function. Currently, there still is a tendency to look to the correlations between single neuron activity and extrinsic phenomena to explain how the brain represents the world.[3] In part this is consistent with a computational conception of brain function, in which discrete structures with precise “memory addresses” comprise the fundamental unit of information storage. Yet critics have long suspected that this attribution of high level function to individual neurons is too simple (Yuste 2015; Brette 2019). It is often caricatured as a *reductio ad absurdum* by imagining a specialized “grandmother neuron” that encodes the appearance of one’s grandmother.

---

[3] Indeed the 2014 Nobel Prize in Medicine was awarded for single neuron recording studies that demonstrated correlations with spatial location.

An alternative to a single neuron approach was provided by the work of the U C Berkeley neuroscientist Walter Freeman beginning in the 1980s (see Freeman and Skarda 1990; and more recently Cossert et al. 2003; and Breakspear 2017). Recording from distributed extra-cellular sources on the olfactory bulb of rabbits, he found that recognition of distinctive odors was associated with distinctive dynamical patterns of activity. In the terminology of dynamical systems theory these patterns of activity are called “attractors” because they are not precise but just highly internally redundant and distinct from from each other and from the more chaotic background from which they emerge. This led him to propose a distributed population-level approach to memory retrieval and perception and to reject single neuron representational approaches.

Although this approach is still not mainstream, this is in part due to the fact that we currently lack technology capable of analyzing the complex attractor dynamics that is generated in networks of thousands of interacting neurons. Recent challenges to what can be described as neuron level function have come from a number of sources. Two kinds of evidence are typical. The first type of evidence comes from long train stimulation studies that suggest that complex distributed and highly coordinated patterns of neural activity can produce complex naturalistic behaviors in experimental animals (see the recent study and review in Halley et al. 2020). This is reminiscent of cortical stimulation effects reported by the neurosurgeon Wilder Penfield in the 1950s (e.g. Penfield and Rasmussen 1950). He reported that stimulation of the exposed cortex of neurosurgery patients could sometimes elicit complex memories and percepts (Penfield 1958). The second type of evidence comes from long term or repeated recordings from single neurons. Contrary to the expectations of neuron level theories, these studies demonstrate that neurons do not retain correlated response tendencies despite stable retention of the function attributed to them (e.g. see the recent paper by Schoonover et al. 2021). This suggests that individual neurons are playing a merely supportive role in a larger collective function, and that their correlations with extrinsic features merely reflect their particular position and contribution to the collective activity.

A musical melody is a useful metaphor for a dynamical attractor. And it provides a useful way to conceive of dynamical iconicity. A musical motif can be recognized by its distinctive harmonic modes and patterns of tonal transitions even though subject to considerable variation, ornamentation, fractionation, and juxtaposition with counter-melodies. A neurological dynamic attractor would be subject to similar variations on an underlying theme and be able to be superimposed with other similar attractors or induce resonant patterns of activity in linked areas. In addition, the presence of ongoing activity would not be that surprising in a system where representations are created and retrieved dynamically via the correlation of spatio-temporal patterns and resonant themes. Ongoing background activity would also facilitate entrainment of similar or complementary attractors that tend to resonate with each other dynamically. Moreover, an attractor forming from “noisy” ongoing activity can be analogized to the way that a complex melody can be formed by resonance with simpler incessantly playing melodies. In contrast, from a neuron-centric conception, ongoing activity is considered noise because it has no recognized function in systems where representations are statically encoded.

Although these considerations do not offer definitive support for a distributed dynamical theory of cortical recognition and memory, they are consistent with it and inconsistent with the neuron-level alternative. From a neurosemiotic point of view, then, distributed dynamical attractors provide ideal models for cortical iconic interpretant production, as well as being suggestive of a resonance-like model for the generation of higher order iconic interpretive processes.

Counter-current information processing

An important property of iconism, that is also consistent with a dynamical attractor perspective, is that inconism is a graded phenomenon. Any two patterns can be judged to be more or less iconic of one another in some respect. This is a critical feature enabling the massive parallelism of brain functions. This distributed parallelism is most evident in large mammal brains.

Consider visual cortical areas. In macaque monkey brains, over a dozen distinct cortical areas have been identified that each appear to be specialized to respond to distinct aspects of visual input. This fractionation of visual analysis enables areas to be specialized to discriminate some distinct class of features, such as colors, contours, topology, relative location, direction of movement, and so forth. Fractionation and parallel processing of visual features is facilitated both by “vertical" input-output connections with subcortical systems and by “horizontal” input-output connections with other cortical areas. Vertical connection differences between cortical areas are reflected in differences in thalamic input. At one extreme is the cortical area that is often (misleadingly) described as “primary visual cortex” because it receives fairly direct, minimally processed inputs related from the retina via the lateral geniculate nucleus. But the majority of functionally distinct visual cortical areas receive relatively indirect preprocessed visual information from midbrain structures. These area-specific functional discontinuities are integrated with one another by virtue of cortico-cortical connection. Each visual cortical area is reciprocally connected with adjacent visual cortical areas. And to make things even more complicated, these cortico-cortical connections are organized in ways that reflect each area’s relative “position” in an iconic differentiation hierarchy.

This connectional asymmetry between cortical areas is reflected in differences in the source and target cortical layers of neuron cell bodies and their projections, respectively (see Figure 2). The basic asymmetric geometry of these connection patterns can be described as linking peripherally specialized areas at one extreme (receiving minimally preprocessed input) to arousal and motivationally specialized areas at the other extreme (receiving input from limbic and para limbic sources. Depending on the species there may be as many as 5 or 6 interconnected areas bridging between these cortical extremes. In general terms, this reflects a flow of sensory information “inward” toward brain systems mediating generalized arousal and an inverse flow of attentional-arousal information “outward” toward peripherally specialized areas (for a recent analysis of this counterstream connectional architecture see Vezoli et al. 2021).

Using a physics metaphor, I describe the central-to-peripheral direction *centrifugal* connections, and the peripheral-to-centrtal direction *centripetal* connections (see Deacon 1989). These two opposed directions of reciprocal connections are also distinguished by their source and target cortical layers. Centrifugal connections originate from neuron cell bodies in layer v (by convention cortical layers are designated using roman numbers i to vi, from surface to depth of cortex) and have outputs terminating in neighboring more peripherally specialized areas in sheet like patterns in layers i and vi. In contrast, in the reverse direction, centripetal connections originate from neuron cell bodies in layer iii and have outputs terminating in neighboring more centrally specialized areas in a columnar pattern in layers iv and iii. This pattern is diagrammed in figure 2.

This pattern, in which multiple cortical areas within a modality span between peripherally specialized and central arousal systems, is characteristic not just of vision, audition, and tactile senses but is also of motor cortical areas. So called "primary" motor cortex is in this respect analogous to a "primary" sensory area, in terms of its extreme centrifugal position and fairly direct connections with peripheral systems. In this respect, this parallelism of connectional organization confronts us with the question: In what way is the generation of a behavior analogous to the generation of a perception? The short answer is that both involve the stage-by-stage differentiation of iconic neural interpretants (generated centrifugally), that become stage-by-stage selectively fine-tuned by peripheral information (flowing centripetally). Elsewhere (Deacon 1989, 2016) I have compared this process to counter-current exchange processes found in a variety of physiological processes from fish gills to bird lungs and engineered systems from multistage flash distillation to multistage ultra low temperature freezers.

In physical systems the countervailing flows are capable of driving exchange processes far beyond equilibrium. So, for example, fish can extract far more oxygen from water than would be possible by passive diffusion to equilibrium (see the right panel of figure 2). By analogy, the cerebral cortex can generate iconic interpretants that are differentiated in stages from central to peripheral areas and fine-tuned moment by moment to adapt to the sensory and behavioral affordances encountered in the environment.

[PLACE FIGURE 2 HERE]

In semiotic terms this complex pattern of interconnected areas generates a simultaneously nested set of iconic interpretants. Each cortical area contributes a parallel more or less differentiated iconic interpretant with respect to the same object of reference. Because they are reciprocally interconnected in constant counter-current relation to each other they will also remain optimally complementary and minimally dissonant with their adjacent neighbors. The result is that many parallel iconic interpretants with different degrees of differentiation are linked and simultaneously superimposed to produce a single holistic percept.

[PLACE FIGURE 3 HERE]

Recursive interpretation

Let's now return to the example of interpreting a windsock as an index of the strength and direction of the wind. As we saw above, this requires the generation of multiple iconic interpretants (evoked memories) and then interpreting a higher order iconic relationship that each shares with the others. In other words, it is a two step two level interpretive process in which the initial interpretants are taken together as a (mental) sign vehicle requiring further interpretation of this implicit iconic relationship between them. To interpret this higher order iconism which will provide the object of the indexical inference a subsequent and separate neurological process must be involved. For this to occur the initial interpretants (memories) need to be differentiated to the point that this non visual feature (the "feeling" of the wind) becomes salient. In this process the memories of past similar experiences which constitute the initial interpretants are likely rendered with minimal differentiation in the visual counter-current hierarchy, probably only to the point where they begin to evoke mnemonic responses in other cortical areas responding to different sensory modalities. This cross-modal "resonance" is probably mediated by long cortico-cortical connections between different modalities at fairly early levels of differentiation.

In the case of the windsock this higher order iconic feature is likely interpreted in a non visual sensory modality (tactile and possibly auditory cortex). This transfer (which need not be across modalities) will also be correlated with a shift of attention, as this new secondary iconicity becomes active and the prior activity fades. This shift of attention is probably additionally facilitated by cortical areas within the prefrontal cortex that enhance or inhibit activity in sensory areas thereby shifting from one modality to another or from one level of iconic differentiation to another.

Symbolic interpretation

Every act of cognition is thus a potentially complex multi step semiotic interpretive process. Even a simple indexical interpretation is the result of a developmental progression of signs interpreting signs. And the more indirect the relation between sign and reference, the more complex the underlying interpretive process must be. This is particularly relevant for understanding the cognitive difficulty underlying the competence to communicate symbolically. This interpretive complexity likely explains why symbolic communication has not evolved in species other than humans, and why it is so difficult for other species to be trained to exhibit even a rudimentary symbolic interpretive competence.

As discussed above, interpreting something iconically is essentially automatic since the only mental interpretant necessary is a pattern of neural activity that is similar in some vague respect. And interpreting something indexically, as in the windsock example, only requires a slightly more complex set of iconic processes operating on lower level iconic processes. In both cases the relevant semiotic affordances are intrinsic sign vehicle properties, just hierarchically embedded. In contrast, the sign vehicles that are used to convey symbolic reference have typically been chosen irrespective of exhibiting any intrinsic iconic or indexical clues to what they refer. As a result, the only cues available to aid interpretation are iconic and indexical relationships between the sign vehicles (e.g, word-word relations) and the conventional interpretive habits shared among speakers of a common language.

So although this same semiotic constructive logic is involved in the interpretation of symbolic reference, it requires interpretant generation at a level higher than indexical interpretation. Symbolic interpretation additionally requires the generation of multiple indexical interpretants and interpreting the higher order relations between them. So just as indexical interpretation requiresThis symbolic dependency on indexicality can be intuitively grasped when pointing behavior helps to fix the reference of a name or declaration, but it is also implicit in the way words in a language effectively point to one another. One could, for example, consider a dictionary or thesaurus to be a kind of vast network of words pointing to other words or phrases, and a sentence as a way for words to constrain how neighboring words are to be interpreted. Indeed, the constructive importance of indexicality for the symbolic interpretation of words in a sentence is exemplified by the centrality of contiguity in syntax, whether expressed in agglutination, inflection, word order restrictions, or the special means required to signal long-distance dependencies.

Ungrounding symbolic reference

A consideration of the complex interpretive infrastructure of symbols is particularly important for explaining why language is so radically different from non-symbolic forms of communication observed in animals and humans. Because symbolic reference is determined irrespective of the physical properties of individual sign vehicles, symbolic sign vehicles can be combined and manipulated with respect to their intrinsic properties without undermining their referential associations. This allows the iconic and indexical relationships between the sign vehicle properties themselves to now become useful. But this poses a problem. How can their referential links to the world outside of symbols be established and maintained despite the lack of clues provided by sign vehicle properties?

In a now famous paper published in 1990 the cognitive scientist Steven Harnad articulated a worry that had long puzzled philosophers of language and cognitive scientists in general. He called it the "symbol grounding problem." The mystery was how arbitrary marks, such as the sounds of speech or the states of a brain, could reliably become correlated with specific referents so that symbolic communication is possible. In other words, without determining this mapping extrinsically, i.e. by using symbolic communication to negotiate the establishment and sharing of these correspondences, how could these mappings ever be established in the first place. If it takes communication with symbols to establish this shared mapping convention between symbols and their referents then we are faced with a vicious regress. In that paper he speculates that symbol grounding must therefore be achieved using non symbolic means.

To understand how symbols are grounded cognitively we need to invert this perspective. We must ask how iconic and indexical forms of communication—which are intrinsically "grounded" due to the sign vehicles sharing properties with their referents—can be used to support representation using ungrounded sign vehicles (aka words/symbols). This is of course the challenge faced by every human toddler. From a semiotic perspective, there is a rich and complex set of social semiotic skills being acquired during the first year of life and significantly prior to the early stages of explicit language acquisition. Seen from this semiotic perspective, then, the explosive growth of language during the second and third years of life is a process in which these earlier iconic and indexical capabilities aid the child's discovery of how to use words and word combinations symbolically (see Deacon 2018).

This is an ungrounding process to the extent that the toddler has to discover how to transfer from using intrinsically grounded to using ungrounded sign vehicles, all the while maintaining referential continuity. This can only be maintained if these iconic and indexical relations are in some way preserved in the transition to symbolic communication. Since properties that could provide referential grounding are absent from linguistic sign vehicles, reference can only be preserved by means extrinsic to them, i.e. in iconic and indexical relations between them (i.e. grammar and syntax).

From index to symbol

Every child navigates this challenge of semiotic ungrounding and eventually replaces interpreting words as indices for interpreting them symbolically. To do this the child must recapitulate the semiotically implicit icon(index(symbol)) interpretive hierarchy. During language acquisition the child begins by acquiring a prelinguistic repertoire of intrinsically grounded iconic and indexical communications (Raczaszek-Leonardi et al. 2018). This will become the scaffold upon which symbolic competence is built. In this process early word uses are indexically correlated with objects of reference by virtue of regular co-occurrence. This is aided by the uniquely evolved human adaptation for pointing to coordinate joint attention. When pointing behavior is correlated with the utterance of a word and linked to a type of object a two-component transitive form of indexicality results. This is critical for learning to differentiate which aspect of a given object is being referred to. Repeated uses of the same word sound are iconic of one another, whereas each instance of its correlation with a physical object or event may be different in many respects. The iconism of a repeated word sound coupled with the expected constant correlation of indexical reference motivates an abductive search for the possibly cryptic feature that is shared in common with the many different objects. In the process the child begins to build a repertoire of indexical associations between words and objects. But this isn't symbolic.

The transition from indexical to symbolic use of words begins as many new words are acquired and the use of word-word relations begins to eclipse their indexical use. This is driven by a mnemonic problem. Indexical associations are one-to-one associations. Because word sounds are conventional the association lacks any iconic support for remembering the association. The only iconism is provided by repeated pairing. But as the size of the word repertoire grows the mnemonic load increases while the regularity of constant sound-object pairing decreases. As is well known, maintaining simple associations depends on reliable stimulus pairing. When this physical co-occurence becomes unreliable (as it does in the case of language use) these learned indexical associations should tend to be extinguished. So if word reference was merely the result of indexical correlation we should expect that word reference would regularly fade. Of course it does not, and even modest vocabularies tend to include tens of thousand words. This is because words are mnemonically supported by their iconic and indexical associations with other words. These associations are stably maintained in the language community and are reflected in their content association in everyday usage in sentences, etc. So unlike simple physical correlations they are not subject to extinction. In one sense, then, the shift to the systemically supported symbolic reference is in part driven by a mnemonic least work principle (though this does not explain why other species exposed to language don't do this).

This is an important clue for understanding the neurological basis for symbolic interpretive processes, and why there aren't new language-specific structures in the human brain. Rather, the distinctive symbolic nature of language has recruited cortical and subcortical systems that evolved in response to otherwise unrelated sensory and behavioral adaptive needs. Moreover, it requires that these otherwise distinctive functional systems work together in novel synergistic ways. This is because of the way symbolic reference is supported by distributed indexical interpretive processes. But in this case, the indexicality is between symbolic sign vehicles (e.g. words) themselves on the basis of their sound iconicity and indirectly by their implicit but displaced referential indexicality. Recall the above description of indexical interpretant generation as second order iconicity facilitated by an attentional shift mediated by prefrontal cortex intervention. So to the extent that symbolic interpretant generation involves second order indexical relationships between sign vehicles that are themselves indexically correlated with their referents there needs to be an iconic representation of this word-word, meaning-meaning network of relations. This necessarily requires a complex functional synergy between frontal and posterior cortical systems.

Interestingly, an early suggestion of this necessary semiotic neurological synergy was provided by the structuralist-linguist Roman Jakobson. In his analysis of sentential structure Jakobson suggested that there were two orthogonal dimensions of sentence processing: which he termed the syntagmatic and paradigmatic dimensions (see Figure 4). Syntagmatic mental operations are diachronic, involving alternating shifts between functionally exclusive sentential units; such as metonymic associations (e.g. tree-squirrel). Whereas paradigmatic mental operations are synchronic, involving simultaneous choice options drawn from within a hierarchically nested set of similar sentential units; such as metaphoric associations (e.g. lion-cat). Notice that syntagmatic-metonymic relations are involve indexical interpretive processes and paradigmatic-metaphoric relations involve iconic interpretive processes. In his brief essay on aphasia (found in chapter 5 of Jakobson 2018) he suggested that posterior cortical regions are critical for the paradigmatic functions, such as choice of the appropriate word in a given sentential position, and that frontal cortical regions are critical for the syntagmatic functions, such as shifting from one word to another different complementarily functioning word. This can be interpreted to suggest that frontal processes enable shifts from one posterior iconic state to another, and that word sounds should be spatially-cortically correlated with the sensorimotor features that serves as mnemonic cues to their referents.

Putting these insights together provides a view of symbolic interpretive processes that effectively involves most of the cerebral cortex, not merely a few "language specialized" cortical areas, as the classic Broca-Wernicke theory would suggest. Indeed recent fMRI research suggests that support for word meaning is distributed throughout the cerebral cortex in both hemispheres (Huth et al. 2016). This suggests that we should begin to think of the generation and interpretation of sentences as a differentiation process (Deacon 2016)—not as a word and phrase combination process—and which thereby takes advantage of the same iconic-indexical differentiation logic as other forms of perception and action. It should not be surprising, then, that imagistic processes are common correlates of the conception and interpretation of sentences.

Mnemonic implications

This further provides insight into a unique mnemonic innovation that symbolic processes have provided to human cognition and which is unavailable to other species (see Deacon 2016). This can be described as narrative memory. In all mammal (and probably most vertebrate) brains mnemonic functions are divided into two quite distinct neurological systems supporting two quite distinct and to some extent functionally incompatible means of storage and recovery (for an overview see Baars and Gage 2007).

The acquisition of a motor skill is accomplished by repetition of an activity, which progressively improves precision and efficiency of the activity. Repetition progressively increases its automaticity and reduces the need for consciously monitoring its production. It is generally assumed that repetition strengthens some complex synaptic pathways and weakens others. Thus retention and recall of this information is facilitated by the way that the signaling has become canalized by redundant synaptic strengthening. This mnemonic strategy is often called *procedural* memory for these reasons. It exemplifies the general principle that mnemonic strength and accessibility is a function of statistical redundancy.

In contrast, it is also critical to be able to store and retrieve experiential information in contexts where repetition is not possible. What has been called *episodic* or declarative memory is memory (I find this latter term misleading because of its linguistic framing) for events or episodes that occurred once, uniquely, are not repeated, or involve little in the way of repeatable actions. This requires the generation of redundancy of a different sort: redundancy of associations. This creates memory traces for singular experiences by correlations between features. Thus when we try to recall a specific experience from our past, it is generally necessary to triangulate to it using correlated associations, involving dates, places, typical social frames, and so forth. This form of mnemonic redundancy is formally orthogonal to the redundancy characteristic of procedural memory.

Because of this mutually incompatible difference in the role played by redundancy, procedural and episodic memory formation are supported by quite distinct neural substrates. Thus, procedural memory is dependent on cerebral cortical (especially frontal) interaction with basal ganglia and cerebellar structures, whereas episodic memory is dependent on cerebral cortical (especially sensory) interaction with the hippocampus and adjacent structures.

Because of this functional segregation of these mnemonic systems, language has come to play an interesting and unique mediating role. It has become the foundation upon which an unprecedented new form of memory has emerged in human evolution. Early in the process of language acquisition articulatory and syntactical combinatorial skills are acquired procedurally, via the many thousands of repetitions of similar word sounds and syntactic habits. In contrast, the symbolic reference that constitutes word meanings and their penumbra of semantic and experiential associations are necessarily acquired episodically in a growing web of associations. Because of this synergistic use of mnemonic systems, language enables each mnemonic system to reciprocally cue the other. In Jakobsonian terms, procedurally acquired syntagmatic associations can cue episodically acquired paradigmatic associations which in turn can cue new syntagmatic associations and so on. This metonymic-metaphoric and indexical-iconic alternation is facilitated by the fact that the sign vehicles (e.g. words and phrases) of language are unencumbered by the need to share intrinsic properties with what they refer to.

Narrative memory is the result. It forms the basis for promising, reasoning, and theorizing, as well as creating our identities, histories, politics, and art. Essentially, every form of socially maintained pragmatic knowledge, from religious belief to technology, is built from a growing matrix of shared narratives of diverse forms. So although the neural substrates supporting these distinct mnemonic systems have not been fundamentally altered in human evolution from the ancestral primate condition, human cognition has been radically restructured by this novel mnemonic affordance made possible by the evolution of our unprecedented symbolic interpretive competence.

Summary and conclusions

This has been a highly speculative effort to find common ground between the neurosciences and a "science" of semiosis. To accomplish this at even a superficial level of detail has required rethinking assumptions that are currently taken for granted in each field.

In the neurosciences this has required abandoning both the computational metaphor and also the view that the basic functional unit of recognition and memory can be identified with neuronal response properties. Instead it suggests that the basic neurosemiotic unit is a dynamical attractor that likely involves the collective activity of thousands of neurons forming a large distributed network within a local cortical area, modulated by its distinctive patterns of input and output relations.

For semiotic theory this has required abandoning a categorical structuralist conception of the sign relation., and instead adopting a process perspective in which iconic, indexical, and symbolic relations are understood as phases in a multi-step interpretive process, and not as sign vehicle properties. This further requires that semiosis is understood as hierarchically constructive with iconic interpretation as most basic interpretive process, indexical interpretation as dependent on prior iconic interpretation processes, and symbolic interpretation as dependent on prior indexical interpretation processes.

The result is a reframing neurological processes as semiological processes and vice versa. From this synthetic perspective the basic neurosemiotic function is understood to be iconic interpretation, whether perceptual, mnemonic, or behavioral. Higher order indexical and symbolic forms of cognitive inference or communication are understood in terms of recursive complexifications of iconic processes. Although this is not sufficient to provide a theory of subjective experience, it does provide the necessary semiotic foundation without which contemporary theories consciousness will remain cryptically dualistic.

Reflecting on where things currently stand, this is at least a beginning. To date the cognitive neurosciences have been almost completely oblivious to the potential relevance of semiotic theory. On the one hand, this is the result of reluctance within the cognitive sciences to address the presumed philosophical challenges posed by the problem of mental representation and instead to frame research in a purely mechanistic paradigm. And on the other hand, it is the result of the failure of semiotic theory to grapple with its physical implementation, and instead to mostly address problems within the humanities where the concept of mental representation can simply be taken for granted.

As a result, for the most part, people working in these fields know little or nothing about current theorizing in the other. This has impeded advances in both fields and has made the problem of explaining conscious experience and agency intractable. It has also led to a resurgence of scholarly interest in the extremes of eliminative materialism and panpsychism. This is not inevitable. It just cannot be overcome without finding a common neurosemiotic methodology.

I have argued that to achieve this it will be necessary to challenge many common assumptions in each field and to embrace Peirce's core insight: that both cognition and communication involve the production of signs; that what goes on in brains is also just the generation of signs to interpret signs, no more and no less. To provide an example of how this might be approached I have described how the hierarchic interpretive dependence of symbols on indices and indices on icons can be understood in terms of neural processes taking place in the cerebral cortex and related brain structures. Of course, only when we are able to trace a specific empirical example of this process in some detail from process to process and structure to structure in the brain will this achieve the status of a semiotic theory of brain function. This may still be a long ways off, but we can't achieve this unless researchers in these currently isolated fields develop some common conceptual tools. I hope to have provided some hints as to how this might be accomplished.

In conclusion, this has necessarily been a highly schematic and abstract account, that will require considerable empirical study to elaborate, clarify, and test. But it offers a basis from which to begin to formulate a new synthesis; a semiotic cognitive neuroscience.

References

Baars, B. J. and N. M. Gage. (2007) *Cognition, Brain, and Consciousness: Introduction to Cognitive Neuroscience*. London: Elsevier Ltd.

Breakspear, M. (2017) Dynamic models of large-scale brain activity. *Nat. Neurosci.* 20, 340–352.

Brette, R. Is coding a relevant metaphor for the brain? *Behav. Brain Sci*. 1–44 (2019).

Copeland, B. J, (2020) "The Church-Turing Thesis", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/church-turing/>.

Cossart, R, Aronov, D & Yuste, R. (2003) Attractor dynamics of network UP states in the neocortex. *Nature* 423, 283–288.

Deacon, T. (1989) Holism and associationism in neuropsychology: An anatomical synthesis. In: E. Perecman (ed.), *Integrating Theory and Practice in Clinical Neuropsychology*. Lawrence Erlbaum & Assoc., Pub., p.1-47.

Deacon, T. (2016) On Human (Symbolic) Nature: How the Word Became Flesh. In *Embodiment in Evolution and Culture.* T. Fuchs & C. Tewes, eds., pp. 129-149 .

Deacon, T. (2016) The emergent process of thinking as reflected in language processing. In- *Thinking Thinking: Practicing radical reflection*. Donata Schoeller & Vera Saller eds., pp. 135-158.

Deacon, T. (2018) Beneath symbols: Convention as a semiotic phenomenon. In Steven C. Hayes & David Sloan Wilson (eds.) *Evolution & Contextual Behavioral Science: A Reunification.* New Harbinger Publications.

Freeman, W. J., and Skarda, C. A. (1990). Representations: Who needs them? In J. L. McGaugh, N. M. Weinberger, & G. Lynch (Eds.), *Brain organization and memory*. New York, NY: Oxford University Press, pp. 375–380.

Gibson, J. J. (1966) *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Harnad, S. (1990) The Symbol Grounding Problem, *Physica D*, v. 42, pp. 335-346.

Hawkins, J. and Blakeslee, S. (2005) *On Intelligence : How a New Understanding of the Brain Will Lead to the Creation of Truly Intelligent Machines*. St. Martin's Press

Huth et al. (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532: 453-458.

Jakobson, R. (2018) Two Aspects Of Language And Two Types Of Aphasic Disturbances (Ch. 5 in) *Studies on Child Language and Aphasia*, Berlin, Boston: De Gruyter Mouton.

Newell, S. W. and Simon, H. A. (1972) *Human Problem Solving*, Prentice Hall, Englewood Cliffs, NJ.

Newell, S. W. and Simon, H. A. (1976) Computer Science as Empirical Inquiry: Symbols and Search, *Communications of the ACM* 19(3): 113–126.

Peirce, C. S. (1931-58) *Collected Papers*, eds. Hartshorne, C, Weiss, P. and Burks, A., Cambridge, MA: Harvard University Press.

Penfield, W. (1958) *The Excitable Cortex in Conscious Man* (The Sherrington Lectures) Liverpool University Press.

Penfield, W. and Rasmussen, T. B. (1950) *The Cerebral Cortex of Man*. The Macmillan Company, New York.

Raczaszek-Leonardi et al. (2018) Language Development From an Ecological Perspective: Ecologically Valid Ways to Abstract Symbols. *Ecological Psychology* 30:39-73.

Schoonover, C.E., Ohashi, S.N., Axel, R. et al. (2021) Representational drift in primary olfactory cortex. *Nature* 594, 541–546.

Vezoli, J, Magrou, L, Goebel, R, et al. (2021) Cortical hierarchy, dual counterstream architecture and the importance of top-down generative networks. *Neuroimage* 225, 117479.

Vilarroya, O. (2017) Neural Representation. A Survey-Based Analysis of the Notion. *Front. Psychol.* 8, 1458.

Ward, LM. *Dynamical Cognitive Science.* (MIT Press, 2002).

Yuste, R. From the neuron doctrine to neural networks. *Nat. Rev. Neurosci*. 16, 487–497 (2015).

Figures and figure legends.

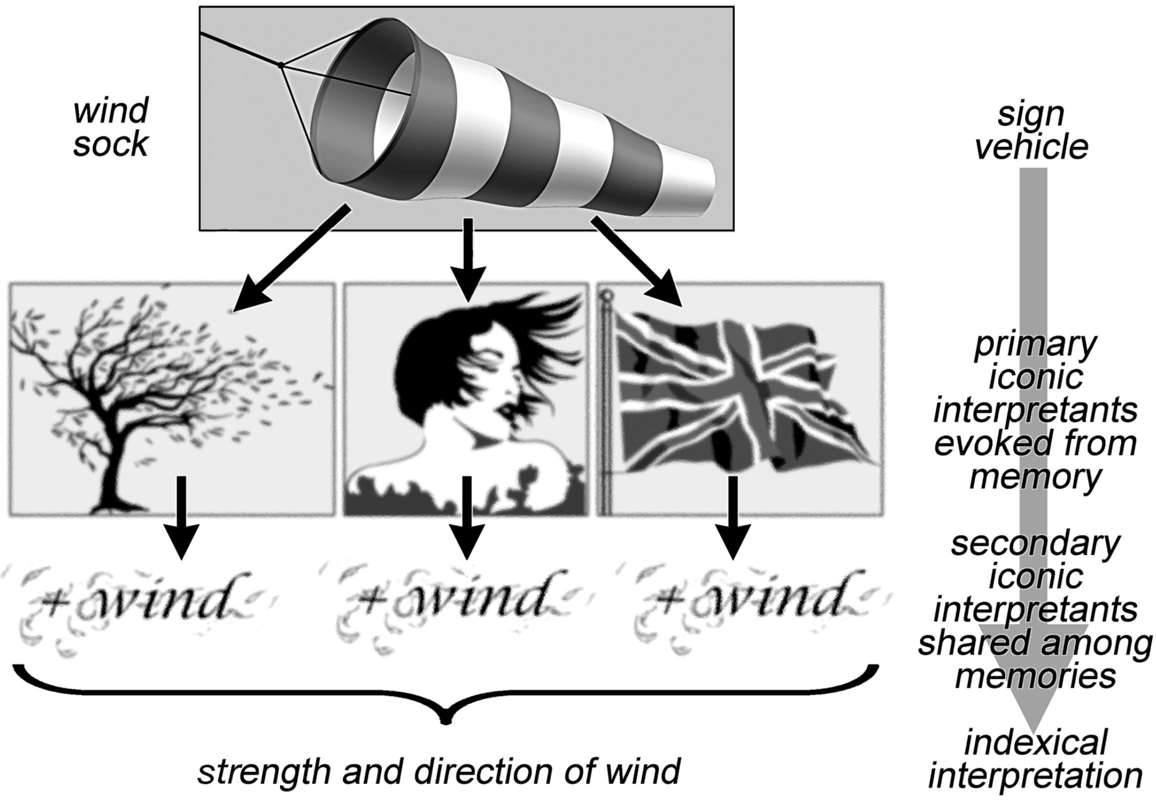

Figure 1. Schematic depiction of the two-phase process of interpreting a windsock to be an index of the strength and direction of wind. A two phase two level interpretant generation process is depicted. The three images shown are intended to represent "mental images" serving as initial iconic interpretants invoked from memory by the sight of the windsock. The second phase is the recognition of an additional iconism linking all these memories that becomes the basis for inferring the unobserved indicated feature: the wind.

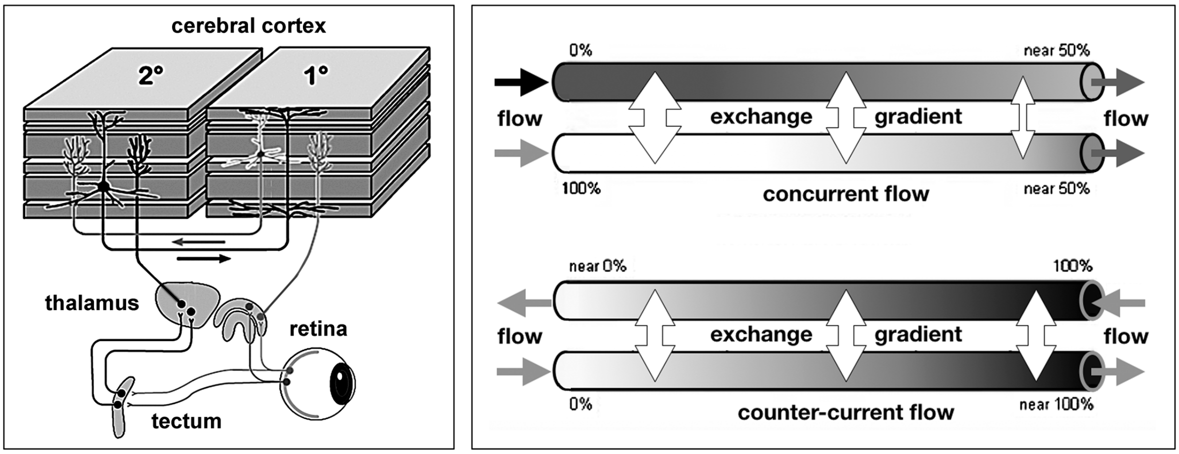

Figure 2. Left panel: Adjacent cortical areas within a modality are connected reciprocally by cortico-cortical connections that originate and terminate in different layers depending on their relative direction in a gradient with respect to how directly connected the area is to peripheral receptors (or effectors). Connections from more peripherally specialized (1°) areas to less peripherally specialized (2°) areas originate from pyramidal neurons in superficial layer 3 and have axons terminating in a columnar pattern within layers 4 and 3. Reciprocal connections tend to originate from pyramidal calls in deeper layer 5 and terminate in a sheet-like patter in layers 1 and 6. Right panel: Two alternative models of dynamical exchange processes, such as in molecular or heat exchange, are shown; generally described as concurrent and counter-current exchange processes. It is hypothesized that the reciprocal connection pattern that links adjacent cortical areas functions analogous to counter-current exchange. In thermodynamic and physiological processes counter-current exchange can drive diffusion processes far beyond equilibrium by virtue of the way it optimizes the gradient between juxtaposed media originating from sources with highly divergent values. Neuro-semiotically this is analogous to optimizing similarity between divergently structured signal patterns (e.g. interacting iconic interpretants). Redrawn from Deacon (2016).

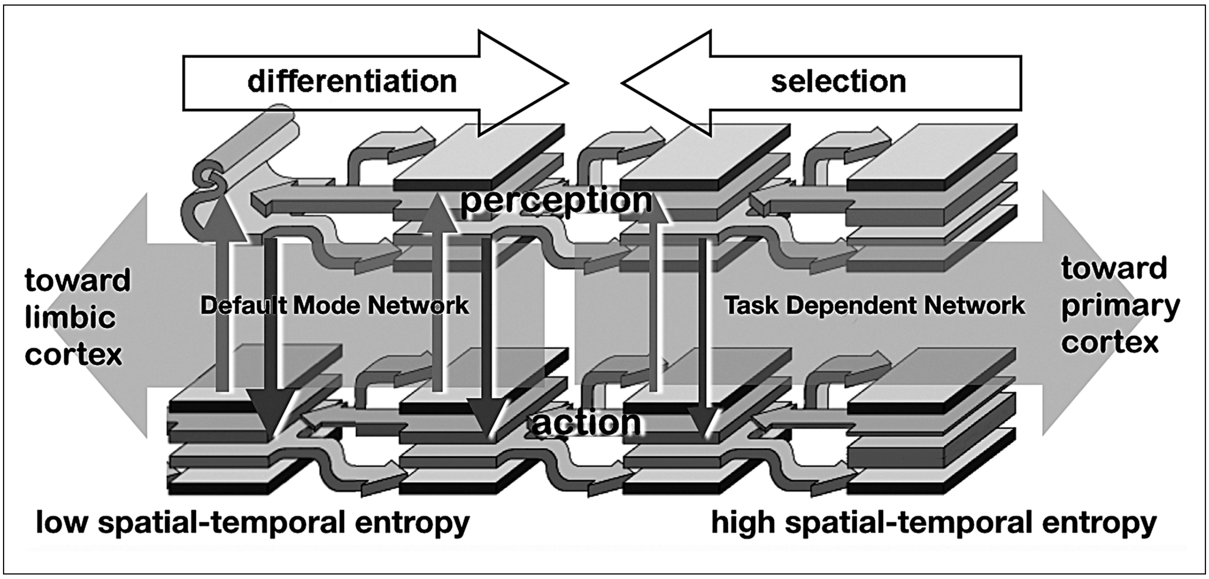

Figure 3. Global counter-current organization of cortico-cortical connections schematically depicting the distinctive laminar asymmetry of connection patterns (see Figure 2 left panel) with respect to the gradient from peripherally driven (toward primary cortex) to arousal driven (toward limbic cortex) areas. These extremes of cortical specialization are also sometimes described as comprising task dependent networks versus default mode cortical networks, respectively. Only one sensory modality is depicted for simplicity. Long projections linking modalities are also depicted (vertical arrows), and tend to be more extensive between areas that are less directly connected to the periphery.

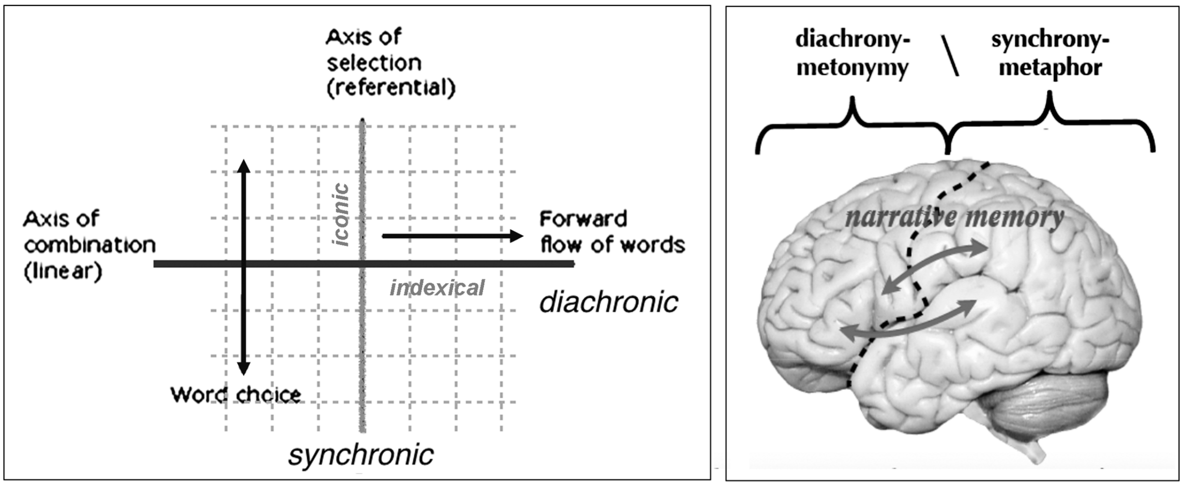

Figure 4. Roman Jakobson's two dimensional analysis of language processes (left panel) and his hypothesis concerning how they are segregated between frontal and posterior regions of the cerebral cortex (right panel). I have linked Jakobson's synchronic-metaphoric dimension with iconic interpretive processes in general and his diachronic-metonymic dimension with indexical interpretive processes in general. As discussed in the text, I also argue that the two orthogonal modes of mnemonic consolidation (described as procedural and episodic memory processes, respectively, and which are supported by distinct neural substrates) are brought into synergistic relation to one another by symbolic language, giving rise to a human-unique mnemonic capacity: narrative memory.