Donald M. O'Malley   Application to National Library of Medicine
© 2010-2011

Neuronal Coding: From Objects to Symbolic Thought   *SAMPLE CHAPTER*

The Labyrinth of Human Intelligence: Many Great Divides
The field of the Neuroscience of Human Intelligence (if it is truly a field at all) is a field of many great divides, with the greatest divide of all being that between single-neuron firing patterns and General Human Intelligence (GHI). The gap between local neuronal circuits and "symbolic thought" is perhaps narrower, but formidable nonetheless. The tools of computational neuroscience (Rieke et al., 1999; Dayan and Abbott, 2005) may seem to take us closer to the true neuronal roots of intelligence, but we will consider if this narrowing of the divide is more illusory than real. A similar kind of divide separates the promises of artificial intelligence (AI) from its best performances to date—an arena viewed as barren by many, but one filled with the best efforts to create new intelligence and a source of many thoughts on just what it means to be intelligent (Legg and Hutter, 2007; Minsky, 1988). Indeed it is difficult even to define intelligence, although we are quite industrious in terms of measuring it. We will defer a definition until later, i.e. after developing some neuronal circuit operations that are likely relevant to a mechanistic discussion of how intelligence might work.

Intelligence has ancient phylogenetic underpinnings. While aspects of intelligence might, in some sense, be emergent properties of complex networks, every emergent property has some pre-emergent basis. To get to the neuronal basis of intelligence, it is best to start small, and then follow intelligence up the evolutionary ladder—starting perhaps with zebrafish, given their object recognition and decision-making capabilities! The course is not strictly linear and the language capabilities of Alex the Parrot (Pepperberg, 2002; 2009), probably do not reflect cognitive capabilities of the last common ancestor we shared with birds some 200 million years ago. But they do tell us what today's parrot brain can do, while the social lives of primates (as in the "non-human" kind, for this work) and hyenas tell us things about what assorted carnivore and omnivore brains can do. The extensive "vocabulary" of Chaser the border collie (NYT-Science Times, 1/18/11), suggests that the ability to acquire a sizable set of learned words (including nouns and verbs) can be found quite widely throughout the avian and mammalian vertebrate lineages. This aspect of animal intelligence, and its neuronal basis, has received only modest attention amongst those trying to understand the heights of human intelligence (and following the lines of Fodor, 1975; Pinker, 1999—see note at very end of bibliography).   There are particularly curious issues with the hominid lineage, and how we went from scavenging carcasses to building space stations. Questions of language and social groupings seem central to the problem of GHI, but what it is about our brains, other than size, that makes us different from Australopithecines is very difficult to answer. Nonetheless, the collected information on animals ranging from fish to hominids, in connection with our ability to probe and understand modern minds, offers a feast of information. Our challenge is to make sense of this, in some integrative way, at the neuronal and local circuit levels.

Neuronal Coding

While the functions of photoreceptors and bipolar cells in the retina are fairly straightforward, as sensory signals propagate deeper into the net (brain), the nature of the neuronal signals becomes increasingly abstract and obscure (Rieke et al., 1999; Dayan and Abbott, 2005). Lower-level sensory features were discovered by accident in neocortex (Hubert and Wiesel, 1959) only because the slide changer used to switch visual stimuli happened by chance to stimulate an orientation-selective visual

neuron whose receptive field happened to be aligned with the edge of the slide changer. While newer neuronal population imaging techniques allow receptive field properties to be studied more extensively and systematically (Niell and Smith, 2005; Li et al., 2008), higher-order receptive field properties may become increasingly arcane and difficult to understand, or even find: the means that we have discovered thus far to drive higher-order neurons are not necessarily emblematic of the essential functions of these neurons.

On the output end, even in the simplest model vertebrate animals, it can be difficult to uncover the logic of the motor control systems. The commands controlling the swimming behaviors of lower vertebrates are complex and at best poorly understood (Zelenin et al., 2001; Gahtan and O'Malley, 2003; Orger et al., 2008; Sankrithi and O'Malley, 2010; Mullins, 2010; Zelenin, 2011). Similarly, the neurodynamics (complete voltage histories of all involved cells) of the simplest of central circuits, the spinal central pattern generator (CPG) is far from being fully understood (Buchanan, 2001; Liao and Fetcho, 2008; Grossman, 2010; Berkowitz, 2010; Dougherty and Kiehn, 2010; Talpalar and Kiehn, 2010). Certainly we know that descending excitation and crossed inhibition are important in driving spinal CPGs and organizing their operations (i.e. to initiate locomotion and to produce rhythmic, alternating bending or stepping), but the number of potential mechanisms compatible with the currently available data is surprisingly diverse and there is no widely accepted answer for how this simplest of central, vertebrate circuits operates. The spinal CPG circuit remains "under-constrained", meaning that we cannot show that there are many possible solutions consistent with the available anatomical and neurodynamic data. If there are many possible answers, then more research remains to be done.

The in-between zone, the circuits between sensory input and motor output, is where neuronal coding get seriously hairy—i.e. vastly interconnected. In mammalian forebrain, the problem is immeasurably complex. That we have stumbled upon ways to drive different cortical neurons should not be confused with neurodynamic explanations. When it comes to the neuronal circuit basis of Daily Memory Records (DMRs), language and intelligence, we have not even begun to understand the potential mechanisms. Because we do not know WHAT these systems are doing, it becomes impossible to figure out HOW they are doing "it".   Nonetheless, there are a variety of constraints that can be brought to bear on these problems, and so it is fair to search for constraints that will narrow the range of potential "solution spaces" within which to search for neuronal mechanisms and corresponding functional capabilities.

From Circuits to Symbols
The problem of neocortical complexity is bad enough given its "rats nest" wiring. [Figure 1. Permission needed from PloS, authors]. It is made perhaps immeasurably worse by neuronal phenotypic diversity. Whether one is talking about the zebrafish brainstem

FIGURE 1. Connections of Monkey Frontal Cortex: PloS Biology, Averbeck and Seo, 2008
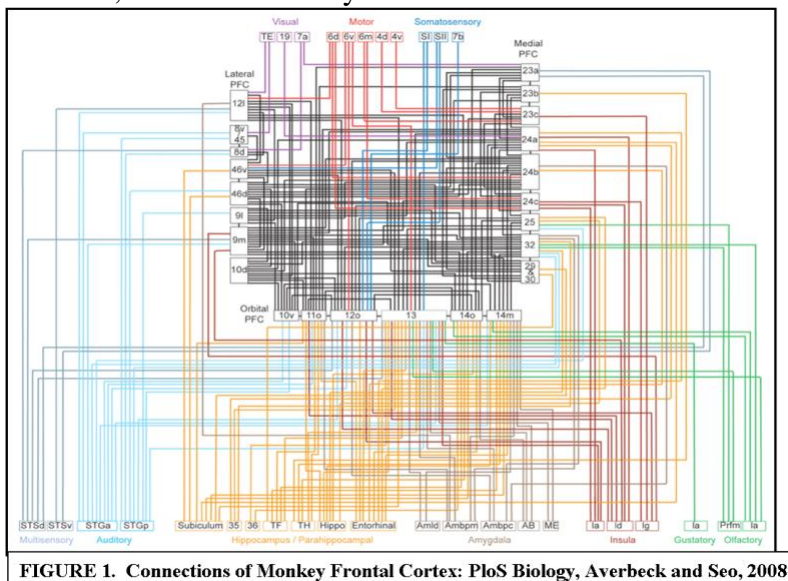
(Niell and Smith, 2005; O'Malley et al., 2003), the rabbit retina (Masland REF) or mammalian hippocampus and neocortex (Stevens, Nelson REFs), there is always a tremendous diversity of functionally and anatomically distinct neuronal cell types. This builds in a basal level of computational complexity which is multiplied by the diversity of connection patterns—embodied by

the newer field of connectomics (Seung, 2009).   This increases the number of possible computations many fold. How then can we constrain our thinking about this problem, i.e. our theoretical analysis? In this work, this effort will be made based upon two principles:

(1) we will try to understand a core set of functional capabilities or "neural operations" that fits with and seems important for observed behaviors and behavioral capabilities, and

(2) we will further constrain the plausible solution spaces by taking into account neuroanatomical and biophysical constrains (which can also be viewed as computational opportunities).

These two principles will be employed in an effort to get at the "local-circuit" basis of such cortical operations as language, intelligence and memory, including our day-long flash memories, i.e. our Daily Memory Records (DMRs). Of most immediate concern is the generation of neural "symbols" that our brains can use to adaptive advantage. Such symbols must of necessity be built upon more primary functions such as object recognition and invariant representation. These two basal functions lead to the concept of "Neural Words", that can be tied into key and perhaps universal cortical operations, and that may also constitute a neuronal nexus of symbolic information processing.

From Neuronal Networks to Object Recognition

A "form" of **Object Recognition** is evident in one of the simplest of vertebrate animals: the larval zebrafish. Without any prior experience, a 6-day old larval zebrafish is able to identify a small moving object, such as a paramecium, as a potential food source. This is indicated by their innate ability to perform a sequence of complex motor patterns (J-turns, tracking swims and capture swims) leading to successful prey strikes the first time they see a paramecium in their vicinity (Borla et al., 2002; McElligott and O'Malley, 2005). Zebrafish are unfamiliar with "phylogenies" or the concept of microorganism, but so of course are all other animals, excepting man: many animals are capable of recognizing objects of myriad types, regardless of whether or not they think of them as objects. Human babies quickly begin exploring their world and learning about objects. One difference is that human babies are helpless, while larval zebrafish are sufficiently good predators as to survive entirely without help—for long enough to mature and propagate their species.   There are crucial differences between innate (genetically-encoded) knowledge of objects vs. exploratory learning about them, and both learning and genetic-encoding of world knowledge will loom large in our investigation of GHI, but for now these two behaviors have something in common: objects are things of potential relevance that must in some sense be understood. Babies explore with their eyes, hands and especially their mouth (which is why one should not give them marbles, bottle caps, etc.).

A larval zebrafish might be said to "explore" in that it will track and approach a prey item, e.g. a brine shrimp, but then abandon its attack if the item is a too large to be engulfed (R. Westphal, unpublished data). It is hard to know what is in the mind of the larval zebrafish (or even if it is conscious), but we expect that during a bout of prey-tracking and prey-capture, the "identity" of the object is not changing on a msec by msec basis as the larval moves closer to engulfing the item. The apparent size of the object (the size of the paramecium's image on the larval retina) increases continuously up to the moment of the strike, but throughout the attack the larva behaves as if this is an item that should be eaten (if small enough). Thus, this larval zebrafish might be said to have an "invariant representation" of the object being tracked, even if it did not know the object's precise size when tracking began. While some distant ancestor of the zebrafish did NOT know that microorganism-sized, moving objects would be good to eat, we can presume that over evolutionary time the motivation to approach and engulf small moving things became genetically engrained, and that this was accompanied by an ability to ignore other objects, while fleeing still other objects.

3

The zebrafish behaviors can be described as reflexive, as long as one assumes these larvae are unconscious and lack free will, but these larvae certainly possess neuronal circuits that choose amongst at least 3 possible options: ignore, approach or flee objects in the world—and so larvae must continuously scan their environment and categorize the world in some such terms. This larval decision-making process can be described as a winner-take-all (WTA) algorithm. For example, at each stage during a predatory episode, the larva will do either a J-turn or a capture swim, but not some composite of both. Similarly, if a larger predator appears along with the paramecia, the larval zebrafish will pick one behavior to do (e.g. escape, or remain motionless), but it will not do a composite of e.g. escaping and prey tracking. In a simpler scenario, where many paramecia are present in a strikeable range, only one paramecium will be tracked—the larva "pays attention" to just one object at a time! While the number of labs doing neural systems research on larval zebrafish has grown steadily over the past 15 years, the neural circuit bases of the larva's object-recognition and winner-take-all decision-making capabilities remain completely unknown. But the number of neurons involved is small, in relation to mammalian networks, given that the entire larval CNS is (very roughly) 150,000 neurons in total. The entire CNS of the larva is thus on the order of the size of a single columnar structure of mammalian neocortex, e.g. a rat "whisker barrel" (or a small number of such whisker barrels). But what we get out of this small CNS is object recognition, invariant representation, decision-making, flexible behavior and WTA algorithms, along with many other neural functions. This is thus a neuronal starting point for hunting bigger game.

From Zebrafish to Humans. In the case of the human infant, we know that they begin exploring their world very soon after birth (or perhaps earlier in auditory and other senses). Their object-encoding capabilities exceed those of the larval zebrafish by roughly a million-fold, just in terms of numbers of neurons involved. What does this numerical advantage grant humans (young and old) in terms of "object intelligence"? Despite the incredible 3-step tool use demonstrated recently by crows (PBS-Nature, 2010; and see Weir et al., 2002), at some time over during the last 2 million years humans made a substantial leap (perhaps in small gradual steps) from finding things in their environment to manipulating them—in far more complex and flexible ways than beavers building dams or bald eagles building nests. This capability coincides with the much longer post-natal development period of humans, a luxury afforded by the increasingly complex social structures of hominid groups. Our years long (infant-toddler-small child) vulnerability is a trade off for the vast amount of continued neural development that is ongoing throughout, which moreover is multiplexed with ongoing learning—learning associated with both a steady, unavoidable stream of incoming sensory experiences, and with the more active or cognitive exploration of the world and its objects. Some "simpler" aspects of our postnatal development and learning processes have been demonstrated as e.g. the experience-dependent development of ocular-dominance columns in our visual cortex (REF), and the emergence of orientation columns in ferrets as revealed by neuronal population imaging (Li et al., 2008). Stephen Grossberg has worked out models of how visual categorization and subcategorization can be derived from known features of the visual system's functional and anatomical architecture (1988), but we are still far from understanding what is happening in the infant-to-toddler transition, by which time human object mastery (recognition, utilization, naming) has far exceeded anything any other animal can do. Could larval zebrafish neural circuits do these things, by simple replication—just copying the circuits many times and hooking them in to the right sensory inputs? Or are there neuronal phenotype-specific functions that enable infant neocortical pyramidal cells (and other cortical/hippocampal/cerebellar cell types) to do things that zebrafish neurons simply cannot do, regardless of simple replication? A third alternative is that our *object intelligence* arises as an emergent property of subtle features of very large scale integrated neuronal circuits, and that this imbues us with capabilities not attainable simply by replicating the zebrafish CNS a million-fold and daisy-chaining these zf-modules together.

Computational Implications: While the artificial neural network community can model WTA processes using a simple type of "neuron" called an "integrate and fire neuron" (REF), the behaviors of zebrafish neurons seem more complex in terms of the way they integrate information from different sources and in terms of how they contribute to downstream information processing (Faber et al., 1989). But perhaps these anatomically diverse zebrafish neurons are essentially just "integrating and firing" like the artificial network neurons? It is possible that all of the neuronal circuit complexity needed to explain our "object intelligence" can be generated by a homogeneous neuronal population simply based upon their patterns of synaptic connections and synaptic strengths (weights), which would be adjusted during learning by a unitary Hebbian (or backpropagation) learning rule (Xie and Seung, 2003), to store the newly learned information. We might explain the diversity of neuronal shapes seen in every neurobiology textbook as just a means of ensuring that the proper numbers and patterns of connections could be made to and from each neuron. But there is also a diversity of ion channels, including dozens of sodium, calcium and chloride channels and more than 100 types of potassium channel (Purves et al., 2007). This diversity of channel types is reflected in a diversity of neuronal types, in terms of their electrical activity – their action potential firing patterns, activation threshold and other parameters. Yet a dynamical systems approach (Izhikevich, 2006) suggests that there are only 4 discrete types of dynamical behavior into which all neurons can be grouped, a view echoed by the discrete time-map neural network modeling approach (Courbage et al., 2007), which has been successfully used to build neural-network controllers for robotic lobsters (Ayers and Witting, 2007). Thus, the extent to which neuronal phenotypic diversity (NPD) might be essential to the evolution of custom neural solutions to real-world problems hangs in the balance. We will deal with the more general issue of learning, and an associated diversity of learning rules (LTP, STDP, neural oscillators, etc.), in a separate chapter, but it must be considered that variations in learning (neural plasticity) rules, are going to yield fundamentally distinct computational units, particularly when used in conjunction with differential efficacies of excitation or inhibition along a cell's dendritic tree, with distinct current time courses and hence resonances, both cellular and local-circuit based. The issue of neuronal diversity might seem an obvious one, but it has been largely glossed over (or omitted entirely) in the artificial neural network literature, and in the discussion of human cognitive functions as well. We will return later to the issues of NPD and network construction, but we proceed next to the functional considerations associated with the creation of "Neural Words".

The recognition (and understanding) of objects, in conjunction with spatial memories, comprises a big chunk of what is out there in the world (especially if one adds in a temporal dimension, along with the behaviors of living objects). Humans are extremely good at remembering all the objects they saw in any given day – being able to recall well into the thousands, as demonstrated by Aude Oliva and her colleagues at MIT (Brady et al., 2008). Over the course of the day, humans effortlessly store such object information into their Daily Memory Records, along with the spatial, temporal and other context associated with each object. To store ALL of the information associated with our day's experiences from scratch (objects, world map, behaviors of living things) would be a computational nightmare, not to mention a memory hog, and so we need a shorthand method to encode the information, after which we can effectively and efficiently use it. Object recognition provides not only a means to store incoming information, but also allows us to use our cumulative, stored information about objects (red apples are good to eat, bees sting, etc.), by perhaps relying on an autoassociative neural network that can take an incomplete or noisy data set and retrieve a stored and idealized version. The higher-level *invariant representation* of objects is a more sophisticated analysis wherein we recognize and categorize objects regardless of such variables as distance (within limits), orientation, lighting, and completeness (our zebrafish example being a rather pale version of primate skills in this area). This idea extends to many different arenas, as e.g. our ability to recognize a simple melody regardless of whether it is sung, hummed, played on a violin, piano or kazoo. There is a fundamental representation of the tune that is activated independently of the kind of sound being used to create the musical

sequence. In terms of information compression, it may be that once some physical object is recognized as being present at our location, no more cognitive effort or neuronal resources are required to compute that recognition-- and so one can imagine that the brain might represent the object's continued presence in some more economical form. This is the basis of the "neural word" concept: a neuronal symbol that might serve as a precursor to actual linguistic tags, i.e. real words.   The concept of neural words stems from a work by Jeff Hawkins (2004) laying out a general function of neocortex: prediction. After considering in more anatomical detail how objects are represented in our brains, we return the highly complementary concepts of neural words and cortical prediction.

Neuronal Connectivity & Object Recognition. How do we compress a world's worth of incoming information into a suitable form for object recognition, and subsequent symbolic manipulation? In the case of the larval zebrafish, there is no "learning" component to the first predatory strikes—they use a well-formed, highly-effective strike behavior beginning with their very first attempt (Borla et al., 2002). Motion sensitive cells tell the larva's main visual center, the optic tectum (the piscine homologue of the mammalian superior colliculus) that motion has occurred, while direction-selective cells plot motion trajectory. As the larva moves closer to the paramecium, the paramecium's apparent size will increase, requiring that the object recognition be somewhat size-independent, even though its small initial size is likely central to eliciting the tracking phase of the behavior (any Engert Lab REF?). Some optic tectum cells described by Niell and Smith (2005) have receptive fields that might individually or collectively allow responsiveness selective enough to recognize the key features:

$$small + moving = potential\ prey\ item$$

and larval vision is probably augmented by other senses at the moment of the strike (R. Westphal, unpublished data), but we do not know yet if any of these optic tectum nerve cells are actually involved with prey tracking and capture, i.e. we do not know if they directly contribute to the J-turns, tracking swims or capture swim. Nonetheless, innate mechanisms are able to create paramecium detectors and use them to drive sophisticated motor programs. This might at first blush seem an ultimate compression of information, where a single nerve cell is able to recognize a paramecium, but even in this example, there is a kind of "distributed" recognition system, in that detection units must be spread across the optic tectum so that paramecia can continue to be recognized at successive visual locations as the larva is lining up for the final strike (McElligott and O'Malley, 2005).   As larvae mature into juveniles and then adults, their preferred prey items increase in size (and mobility) and so the triggers for the prey capture behavior (which is also maturing; R. Westphal, unpublished data) are changing, most probably via modification of the innate object recognition subroutine. Zebrafish also recognize predators, conspecifics, container walls and movement of their environment (i.e. water currents) and respond to these items as well. Not a huge representational load, but they presumably use only a small fraction of their ~ 150,000 neurons for this specific task. The object recognition routine is also not very specific in that it won't distinguish paramecium from similar-sized, moving micro-organisms, and the recognition function is entangled with a tracking function that drives brief turning and swimming bouts to bring the item into strike position. But how does this compare with what babies do?

A rattle is pleasing because it makes an interesting noise, is easy to hold and shake, does not taste like lead and is not easily choked on. Babies prefer to look at novel objects, so at an early age attention is engaged by some combination of novelty and salience. I don't know how quickly babies add to their repertoire of remembered objects, but their object recognition skills presumably involve millions of neurons, once one takes into account primary visual and auditory cortex and a plethora of downstream, higher-level and associational cortices. Most of these neurons are connected to thousands of other neurons, making the number of possible paths of information flow essentially infinite: how do we understand the neurodynamics of such a system? [*Neurodynamics* refers to the electrical or voltage histories of all the neuronal units participating in the operation at hand]. Visual, auditory and tactile

features of the rattle must all be processed and categorized and then bound together in some enduring way so that we know (a) what a rattle is, distinguishing it from other objects, and (b) so we can subcategorize it, as e.g. my rattle, my favorite soft rattle, a hard-rattle with an unusual sound, etc. But how are all the attributes of ONE object (e.g. my favorite rattle) bound together in the first place? To be able to recognize an object means that it has features A, B, C and not X, Y or Z. How is this accomplished?

The thousand-fold (or up to 10,000-fold) connectivity of the individual neurons within our neocortical networks enables every neuron to be connected to every other neuron by no more than 3 or 4 synapses, in a simple analysis. Thus neuron #1 is connected to 1000 others, those 1000 to 1 million other neurons, and the 1 million to 1 billion—in a simplistic model with an exclusively branching connectivity (meaning that we are ignoring for now all cortical system topography and many other neuroanatomical details). In some neocortical (e.g. perisylvian) and hippocampal (CA3) structures we do find a pattern of neuronal interconnectivity along a related line, namely that of an autoassociative network (Pulvermüller and Knoblauch, 2009; Lisman CA3 capacity REF). Baby Jane, in learning about her favorite "moo-moo" rattle, might have a cell in her association cortex that fires (in one scenario) when a quasi-rhythmic sound dominated by certain frequencies co-occurs with the movement of a small cow-head with a black-and-white cowhide pattern, held by a smooth, hard semi-circular handle. One problem is that a "smooth surface" neuron in the hand region of primary somatosensory (S1) cortex might send axons to only 1000 or so other neurons outside of S1, and presumably only a small fraction of these project to the ventral visual stream where visual objects are encoded. A similar constraint applies to the neurons in primary auditory cortex that respond to the set of frequencies that uniquely describes the sound of the moo-moo rattle. If we could get these neurons, along with the ones in visual object areas that recognize the cow features of the moo-moo rattle to all project directly to the key association cells, we would now have a few neurons, or even a single neuron that responds specifically to MY "favorite" rattle. [This is more than an academic question, because the loss of such linkages may be at the heart of the memory decline in Alzheimer's disease]. Baby Jane, in initially learning about her "favorite rattle", is able to associate these diverse stimuli (in theory, in this simplistic scenario), because the auditory, visual and tactile information converge upon this lone neuron. It is the synchrony of these inputs that makes the stimulation strong enough to bring the cell to its action-potential firing threshold, which in turn triggers an increase in synaptic strength specifically for the co-activated synapses. Thus, the coincidence (synchrony) of firing of the auditory, visual and tactile inputs converts this cell (or small set of cells) into one that remembers these inputs and responds more strongly to it. This can further result, by the synaptic enhancement of reciprocal interconnections, the ability to retrieve the entire object representation when starting with only partial information, e.g. if starting with just the sound of this particular rattle. We will address later the issue of *inductive bias* (Baum, 2004), which likely predisposes us to learning (more rapidly) certain kinds of things, but for present, we are concerned with the coding of information about this new object in our world.

One can imagine that a small number of cells firing, possibly even one cell (as with the iconic "grandmother cell", although this is taking the idea to an unrealistic extreme; Bowers, 2009), would be sufficient to inform the rest of the brain that I have my favorite moo-moo rattle! Of course, in this minimalist depiction we have ignored a bevy of issues. From an organizational and cognitive perspective, a key omission is categorization: my "favorite rattle" is a subset of the *Rattle* category, and we have ignored that aspect. Also, the cortical architecture and neurodynamic description is without detail: real-world neocortex is six-layered with intra- and extra-column connections, as well as defined topographies of connections to distant cortical and subcortical regions, not to mention recurrent feedback loops. We have not explained contributions of different neuronal phenotypes, cortical up and down states, oscillatory networks nor, social-emotional ramifications of the rattle. But, for present purposes, the most important issue is that we need the brain to produce an Invariant

Representation of the object "rattle", and quite recursively, an inner Invariant Representation that distinguishes our favorite moo-moo rattle from all the others, which pale imitations are clearly unacceptable to Baby Jane.

Invariant Representation.

A central controversy in systems neuroscience is the degree to which representations of things (objects, people, concepts, etc.) are either widely distributed across a neuronal population (fitting with the so called PDP or parallel distributed processing model), or more discrete as in the case of the iconic grandmother cell (Bowers, 2009).   Single-neuron recordings from the medial temporal lobe of humans, including the hippocampus and adjacent neocortex (Quiroga et al., 2005; 2008), reveals surprising response-specificity of neurons to e.g. images of specific buildings or individuals, and moreover shows that these neurons are able to "recognize" (respond to) their preferred objects from many distinct views and even in response to the written name of the object. While Bowers (2009) has interpreted this to support a grandmother-cell like coding scheme, where every object in the world has some few cells dedicated to representing it, Plaut and McClellan (2010) have countered this view, arguing that PDP models better describe the neuronal representation. Quian Quiroga and Kreiman (2010) take an intermediate view, emphasizing the "sparseness" of coding of higher-level constructs, while assigning a more conventional PDP architecture to the task of extracting lower-order features, such as edges in a visual scene. In their view, at higher levels of processing, the lower-level features are successively assembled to build more "grandmother-cell like" representations.

There are some apparent limitations to this analysis, the first being that these cortical recordings are sampling only a tiny fraction of the local neuronal population, and this sampling is biased towards larger neurons since they are more easily detected by extracellular recording electrodes (Logothetis REF). Equally problematic is that we do not know HOW the local neuronal circuits are performing their natural functions, nor even what those functions are (we presume they are somehow involved in functions associated with individuals, buildings or grandmothers, but absent knowing their function, this is only a presumption). The root problem is that we do not have a broad enough or detailed enough view of the network to say much about the computation being performed. Human brain mapping techniques (PET, fMRI) tell us which brain regions show enhanced blood flow during different experimental tasks, but do not tell us what any specific nerve cells, or specific classes of nerve cells, are doing during the tasks.   What is required are recordings of the firing patterns of extended neuronal populations, which can now be done in mouse and ferret neocortex using calcium imaging techniques (Kerr et al. 2005; Li et al., 2008), but this approach is still in its infancy and does not currently seem suitable for studies in primates or humans.

Controversies and technical limitations notwithstanding, we must still posit neuronal architectures that get us to an invariant representation if we are to think about human intelligence. Autoassociative networks (Hertz et al., 1991) are a good starting place because they have the key capability of completing a pattern if provided a sufficient starting pattern. While this would be good at classifying rattle-like objects as rattles, it could also be used to distinguish the moo-moo rattle from inferior brands by virtue of its distinguishing features. The CA3 region of the hippocampus has anatomical characteristics of an autoassociative network, and is strongly associated with spatial maps of the rat's world, and so is one place where we might expect to see pattern completion (where, e.g., seeing just the handle of the rattle helps the baby to find it).   As adults we have more powerful pattern completion skills that includes both subconscious expectations, as well as more attentive and deliberative processes that focus on a specific item and its nature, but early on, we can imagine a world with many fewer experiences, and over time as Baby Jane uses the rattle, drops it, throws it and has it handed back to her countless times, every experience is new, but also overlapping with old versions and the

different views and aspects of the rattle successively build up their connections to a "master representation" of the moo-moo rattle.

Hyper-Associative Memories. Baby Jane's experiences link the representation of "rattle" with her flash memory system, and although we don't yet know the rate or extent to which infants develop full-blown daily memory records, once they have object permanence, they are at least a step closer. If we assume that Baby Jane's DMRs work to some extent like ours: being vast on the day of writing and disappearing over the next several days, this allows her discrete episodic experiences with the rattle to be saved each day, and to the extent that we reinforce those episodes on subsequent days, we enhance her flash memories, converting them into enduring episodic memories (at least up until childhood amnesia kicks in and erases them). In a controlled environment, to test representational invariance, we might present 3 different views of a rattle to Baby Jane—as is often done in testing the ability of machine learning algorithms to solve such problems. But in her crib, Baby Jane sees hundreds of different views, actions and contexts all associated with her moo-moo rattle, and she is not rattled by this at all! The problem of building up an invariant representation is not that of taking 3 exclusive views and fusing them to one object: the problem is one of taking myriad experiences with the rattle, sorting them into different ways of viewing (seeing, hearing, touching) the rattle, which might e.g. include a view where it is hanging on a hook out of reach, a view of it in Mom's hand coming towards me, a quite similar but emotionally different experience of it being taken away by Mom's hand, or a peckish view of it, after having thrown it to the floor. Each experience can be considered to represent the rattle in a high-dimensional autoassociative network, but each is different—each involves a different population of nerve cells. The computational problem is to make sure that this extended "set of networks" are all reliably able to trigger the master "moo-moo" network, if indeed Baby Jane has an invariant representation of her favorite rattle (which it seems she does, given her rejection of poor substitutes).

In one view, over time we acquire a range of possible representations of the moo-moo rattle, aided by cortical predictive networks (more on this later), and perhaps each of these in turn can trigger a master "invariant representation" of the moo-moo rattle. Or perhaps there is no need for the master version— every member network of the moo-moo rattle set is equally capable of triggering the comfortable feeling and joy of this favored object. And why not? To this point we have focused on many-to-one computations (many views, one rattle), but we have ignored the inverse, one-to-many problem: the rattle is many things. It is something you can chew, shake or throw. It is something people can give you or take away. It is something you can see and hear. It provides a way to get people's attention. It is fun. Each aspects of rattle ties it to many other things—you have a network that categorizes things that can be chewed, and the rattle is part of it (maybe not recently, but still). Quite a few things can make noise when shaken. And beyond categories, will have yet another world of associations linked to rattle: the crib, the mobile floating above it, standing up in the crib to reach for the rattle, the baby's room, the stroller, the people Baby Jane shares her rattle with, and so on. The thousand-fold connectivity of each cortical neuron entails something more than an autoassociative network: it entails our starting point for thinking about invariant representations and the multitude of things each such representation is connected to. When I first read Brodal's Neurological Anatomy (1981) (under the kind tutelage of Sanford Palay), it seemed that every part of the brain I read about was connected to every other part leaving me in a state of confused bewilderment. But our hyper-associative memory capability demands this—it demands that every possible thing in our world, every object, place, feeling, plan and idea, be linkable to EVERY other one. Intelligence is about what we do with this cortical miracle.

In trying to understand object recognition and memory, we have encountered a problem—the problem of doing calculus at 10,000-fold connectivity raised to the 1000th power. This is a space in which the

number of potential synaptic pathways becomes essentially infinite, and so how do we navigate this space? Like everything else in Biology, nothing makes sense except in the light of evolution, and so we should turn back to that great instructor, the one that made newly hatched larval zebrafish into consummate predators. Such capabilities evolve gradually, one or a few steps at a time, as e.g. when Home erectus diverged from Homo habilis, perhaps by a slight change in foraging strategy (Bickerton, 2009, p. 150), which then evolved into the ability to organize flash mobs of stone-throwing carcass thieves, in one author's view. The conjunction of within-life experience and a genome geared up to optimize experience-guided behavior is the foundation for the amplified learning seen in many vertebrate animals (Baum, 2004), and this is something we will come back to a bit further on, but we should first try to delimit and reduce the scope of the **object recognition** problem.

Information Compression, Neural Words and Symbolic Expression

Outlined above we see the brain doing PRECISELY the opposite of what we might like it to do: we see the brain making mountains out of mole hills, or moles or moo-moo rattles. I think that instead of taking large, distributed representations and serially multiplying them (like both anatomy and functionality suggest is happening), we need to excerpt the most essential core attributes of objects (and other things) perhaps along the lines of the "principle component analysis" mathematical technique that extracts major features from a dataset. If we can compact the essential attributes into a minimalist neural code, one that is related but not identical to the now popular idea of "sparse" neural coding (Golomb et al., 1990), then we have a tool that can serve multiple purposes (like the hominid stone hand-axe). In other words we need to work towards the physiological suggestions of actual grandmother cells, or at least the "Jennifer Aniston" cells of Quiroga et al. (2005) and for now we will just finesse the issue of 10,000-fold connectivity. But why should we go in this direction? The "compression" of information is inherently good—taking up less memory and less time to transmit, unless of course the receiver cannot decompress the information in a timely manner, or at all. Thus, compression is not entirely win-win, so why should the CNS bother with compression? Beyond the simple neuronal economics of information storage, I view information compression as central to three distinct functions: cortical prediction, DMRs and symbolic thought.

In Jeff Hawkins' influential "On Intelligence" (2004, with Sandra Blakeslee), Hawkins posits ongoing predictions about the world as a unifying function of cerebral cortex—a function that operates whether you are predicting visual or auditory events, movement patterns or the minds of others. This is still, perhaps, the best explanation of why we use the same general 6-layer architecture for every kind of neocortical information processing. This claimed importance of cortical prediction is reinforced by a special theme issue of the Philosophical Transactions of the Royal Society titled *Predictions in the Brain* (Bar, 2009), in which predictive operations were described for a great diversity of neural systems. In their book, Hawkins and Blakeslee link two key observations. First, behaviorally, we (usually) respond quite quickly to anything amiss, e.g. a stranger walking into our house unannounced or a student suddenly climbing up onto his desk and singing the national anthem. One explanation for the immediacy of our response is that the cerebral cortex is a continuous-prediction machine, taking in the world, moment-by-moment, and forecasting, in real time and based upon past experiences expected imminent changes in our environment, as well as expected stabilities. For example, our desk and chairs should not start moving around of their own accord—that might get your attention pretty quickly! It would only be AFTER you became conscious of this unpredicted movement that your brain would start running down the list of possible explanations: earthquake, terrorist attack, ghosts, aliens, etc. (and simultaneously your brain would be evaluating possible courses of actions—darting into an archway if the whole building is shaking, or perhaps calling ghost-busters, if the furniture movements are more consistent with a poltergeist). This predictive function is matched by a curious neuroanatomical feature: the omnipresent reciprocal connectivity between many different cortical

regions. These connections are viewed as the neuroanatomical basis of cortical prediction and this seems like a perfectly useful thing for cortex to do.

But there is one problem with this posited cortical prediction machine. If we consider the vast connectivity that goes into our "object memories" (all the myriad associations noted above), we run into some difficulty: it seems unlikely that the brain, moment by moment, has time to replay every past experience of every object, person and event going on in the world around us, and yet all of this information is fairly directly connected to the current neuronal representations of the world around us—as noted above for both object recognition and invariant representations. We need some information compression. In Hawkins (2004) view, invariant representations are found throughout cortex (p. 109) and consist of sequences. The coding of sequences is extremely important, but a different aspect is the formation of our enduring memories and the use of these past experiences. The great associativity of memory means that any active representations are going to interact with many different brain regions and not in an orderly hierarchy. What is needed is a compact representation that serves the purpose at hand but does not unduly tie up neuronal (and cognitive) resources. The "Neural Word" is a proposed representation, one that excerpts key components of an invariant representation, but that can reside in a subconscious workspace ready for action, while drawing little ionic power. At this point, it is not clear where or how a Neural Word would exist, and how it would differ from an invariant representation, but it seems quite likely that we need a quite iconic representation of our regular associative memories and we need it for multiple reasons.

If we have a very compact representation of the world around us (Neural Words) that are feeding into the cortical prediction algorithm, that makes Hawkins view of a continuously-running, whole-environment prediction process computationally more feasible, and is consistent with our neuroanatomy and behavior. Moreover, Neural Words help with a second problem: Human Flash Memory, alternatively referred to as our Daily Memory Records (DMRs). You can remember 7 (plus or minus) 2 random words in your short-term memory store, yet recall a day-long record of many different things that you did—as can be seen by simple introspection (Gioioso and O'Malley, 2009). This memory store is written effortlessly each and every day, but depends upon our prior experiences. One means to link the ease-of-writing of DMRs to their dependence on prior experience is to posit that Neural Words derived from past experiences, and representing our anticipated incoming experiences, are resident in the same active subconscious information buffer that serves the cortical prediction engine. According to the cortical prediction theory, while the whole world around us is being continuously and subconsciously scrutinized, only items that meet some threshold of novelty (unexpectedness) and salience appear in our stream of consciousness—and thereby garner extra attention and fuller neural resources (see Baars and Franklin, 2003 for an interesting view of our consciousness-generating mechanism). Because Daily Memory Records are a subset (a small fraction to be sure) of our daily conscious experiences, this means that the Neural Words that went into our prediction operations and thus supported our emerging experience of the world, are now able to do double-duty by assisting in the storage of memory recordings of some subset of those conscious experiences. It seems that the reason our vast DMRs are so effortlessly written is because all the heavy lifting has already been done by cortical prediction and invariant representation: we know where we are, whom we are talking to, the details of our physical space, etc. and it is easy for the brain to draw linkages between these things, highlighting especially salient or novel aspects. While the mechanistic features of DMRs (neural oscillators, silent synapses) are addressed later on, these mechanisms are proposed to interconnect pre-existing Neural Words, along with some new information as needed, and this is what helps us to write these records of our lives.

The Neural Word concept is preliminary at this point, and the above explanation an over-simplification, but there is nonetheless a nicety of its fit to both cortical prediction and DMRs. But where Neural

Words can become truly powerful, and serve as an engine of uniquely human intelligence, is in the realm of symbolic information processing. This is discussed below, but we first consider some issues pertaining to this formulation along with some relationships between our two major symbolic information processors: language and memory.

Behavioral Models and Memories. There is a problem with a neural-word style "compressed" representation in that it does not provide immediate access to our extended associated memories, including the full invariant representation networks, or all of the other associated object-memory networks. We thus have to consider how it can help with the cortical prediction task if it does not have all of the elements that go into prediction. One possible solution is compressed models of the **behavior** of the currently present elements of the world, and these predictive construct are then compared with incoming information. Some savings in terms of "information processing load" is likely garnered by the prior object identification work which "turns on" the invariant representations, which can then in turn activate a neural word (or word-behavior) model. One prediction here is that one should be able to find this neuronal activity (using microelectrodes), but perhaps this process is also indicated in psychological tests by shortened reaction times (REF). Having the neural words for current items in the world "turned on" for cortical prediction purposes, not only allows us to more immediately respond to salient/novel events of any urgency, but the writing of experiences to DMR is enhanced because all of the predictive information is being continuously integrated into an ongoing experience record as we e.g. walk from the train station to our office. Thus predictions and neural words are tightly woven into the spatial and temporal components of our lives—as sequences in Hawkins' view (2004). All of this factors into higher-orders of cognitive activity and intelligence and leads to a view of neocortex as a Big Hippocampus—hippocampus being a classical time and space structure, but one that now, in humans, supports every kind of memory, no matter how abstract (more on the role of space-time maps and cognition later).

The predictive process is thus enhanced by prior experiences—all of the things that objects, people and the world-map have done in the past constitute predictions about what will be occurring in the future (and we will be considering whether or not Bayes Rule is an accurate mathematical descriptor of such predictive processes). The bottom line, as far as the emergence of those items of the world that appear in our stream of consciousness, is that a matching process takes place: anything that matches the cortical predictions (and has no special salience) we never see—these things are safe to ignore. Anything novel, or of particular salience, i.e. a topic of our current motivations or goals, does enter our stream of existence, can be directly attended to, and may become part of our DMR and eventually our enduring episodic or autobiographical memory.

What is the neuronal basis of combinatorial codes?
Why did Hyenas, Chimps, Wolves and Birds never develop an extended, combinatorial language? From what we know of animal communications (REFs), they are more than capable of producing a great many calls that are responded to by their group (as e.g. the many Hyena calls; REF). Most such animals cannot imitate (the term "ape" notwithstanding) and most fail concept-of-self tests (Coolidge and Wynn, 2008; p. 82), although this might instead be viewed as failing a concept-of-mirror test, since the observation is a failure to recognize themselves in mirrors. Such data are taken as evidence against a theory of mind, although food caching birds typically know to move food when a conspecifics observes them hiding it (REF). These are tough problems to crack, and quite often one finds lurking within them the metaphysical specter of Consciousness, an attribute that we cannot provably grant to other animals (including other humans). Similarly, one finds issues of symbolic thought and abstract mental activity to be difficult to attribute to most other animals, but in the case of humans we have no problem with this: we have language.

Origins of Language and Grammar. Human language is a combinatorial construct, where we can assemble 26 letters into an infinite variety of sounds and words, and assemble words into an infinite variety of sentences. As far as we know, there is no other species that can do this. The DMR processing, which is ongoing throughout the day and perhaps based upon Neural Words, may also be a combinatorial process. This raises the possibility that this process might be related to or responsible for the evolution or construction of our combinatorial languages. The subconscious information processing and putative Neural Words that lead to both our conscious experiences and DMRs is combinatorial in the sense that the Neural Words constitute a vast array of items and these have to be flexibly and systematically combined both to generate conscious experiences, and then to process the stored experiences to decide what to consolidate into our permanent memory stores. The excerption process seems combinatorial in that all of the incoming experiences have to be compared with a great many potentially relevant stored experiences (as indicated above under Hyper Associative Memory). That is, we use a combinatorial process to find new associations that are useful—salient, novel or both. Pulvermüller and Knoblauch (2009) discusses auto- and heteroassociative networks in the brain (e.g. perisylvian cortex) and cites examples of sequence detectors and notes that unsupervised Hebbian learning can operate these networks and produce rules of grammar (to make their long paper very short). But I do not think that we need rules of grammar to be produced this way. Alternatively, it is possible, in the universal grammar school of thought, that we already know the rules. In The Talking Ape, Robbins Burling (2007) suggested an answer to this question: verbs and nouns came first, followed quickly by the first associations between them (e.g. Hyena-running). This led to the first rudimentary grammar—with no symbolic-derived grammatical rules of any sort.

Or, to go down a bit of a tangent, one might say analogy is a rule, and that we learned a "verb-noun rule" by analogy based upon real world object-action experiences: fire going-out, deer-bleeding, etc. Analogy is a very potent cognitive force (Baum, 2004), but we have no idea how neurons compute analogy, nor why the same cortical structures/algorithms are used to process every kind of sensory (visual, somatic, auditory), motor, planning and other information in order to produce motor sequence, plans and understand social relationships. Cortical Prediction and Analogy are thus at the **core mystery of intelligence** and both are deeply entangled with memory and language. We focus next on possible neuronal underpinning of language, with the hope that later we can integrate all of these elements into a more comprehensive model of cognition.

Words and ProtoWords. The field of linguistics has perhaps underachieved because of a narrow focus on current human language as opposed to language origins. While one might focus on other animals' "languages" (communication systems), those systems have in fact evolved for far longer than hominid language and despite their alleged simplicity are largely refractory to our understanding (current progress notwithstanding, REFS). As smart as we are, why do we understand so little of animal communications? At least in trying to understand the neuronal basis of human language, we understand our current language and so we can ask how it may have evolved, presumably from some chimp- or other ape-languages, keeping in mind that we currently do not fully understand any other present-day primate communication systems (REFS). While animal communications do offer great insights into many aspects of brain functioning, from an investigative perspective, a theoretical assessment of the evolution of hominid language may provide the best ideas on how modern human language works. There have been quite a few works along this line (see e.g. Bickerton, 2009; Burling, 2007; Diamond, 3rd Chimp; Calvin, 2004; other REFs), but what we need to do is determine how well such evolutionary scenarios fit with a neuronal-level assessment of symbol manipulation.

One issue is whether the emphasis of artificial neural network approaches to linguistics (PDP models) where statistical methods are applied to full-blown modern (e.g. English) language might in some way be misleading (Pulvermüller and Knoblauch; 2009) . The concern is that the numerically-discovered

rules discovered therein, might lead to results that bear little resemblance to the actual evolution of the first "protogrammars" at the dawn of hominid language. This is an important issue because the neuronal bases of the earliest hominid vocabularies and grammars are (at least to some extent) still at work in our brains today. A key issue is thus how fast, and hence how much, natural selection would be able to change the neuroanatomical and neuronal-phenotype mechanisms of symbolic operations over the time frame of hominid language evolution (PNAS ref). There is a great range of speculation on the time frame of language evolution, with Bickerton (2009) suggesting that important changes date back at least one million years ago to the Homo habilis-Homo erectus divide (or transition depending upon point of view), while Julian Jaynes has suggested that consciousness itself appeared only in the last 10,000 years (REF). It is clear that there was a great deal of cultural innovation (associated with e.g. use of fire and distance weapons) during the last 200,000 years, with Tattersall (REF) arguing that symbolic thought emerged quite recently, well after the speciation of Homo sapiens (perhaps in the last 50,000 years) and not as a direct result of natural selection. At the other end of the spectrum, our australopithecine ancestors of 4 million years ago had clearly diverged from the remainder of the primate lineage and while their brains were about the size of chimp brains, important neuronal-level changes may have been in the works for at least that long. In the face of this uncertainty, it becomes important to ask what kinds structural changes (anatomical, neuronal diversity) might have been taking place during the evolution of hominid language and exactly how much evolutionary time was there for natural selection to shape these changes?

Early hominid word usage likely evolved over some substantial amount of evolutionary time (given the computational and neuroanatomical complexities discussed herein), while grammars probably began emerging from very simple "rules", e.g. rules governing two-word sequences. But these "rules" were presumably just learned physical relationships, based on stored memories, as opposed to grammatical rules. Because these rule were likely developing over a hundred thousand years or a million years or more [to be expanded upon later], it seems there was a significant amount of time for evolution to select for progeny who were better at understanding physical relationships and events, remembering them and communicating them to other. While this does not rise to the level of (nor necessarily support) the ideal of a universal grammar, it does fit with the more general idea of inductive bias (Baum 2004), which argues that we are genetically endowed with the ability to learn certain kinds of things faster by the genetic restriction of our search spaces (which is analogous to the operational design of how the world chess champion, Deep Blue, tailors its exploration of future chess moves).

In regards to the evolution of grammar, Burling (2007) suggested that protogrammars evolved as a direct result of experience (i.e. memory) (also see Calvin, 2004). Verbs and nouns came first— probably quite a few of them. While Pulvermüller and Knoblauch's associative nets (2009) might in theory play a role in discovering lexical relationships, the experiences that hominid brains needed to understand included myriad physical relationships: hyena-danger, stone-cuts, meat-filling, etc. [While early hominids presumably did not use words as conceptually rich as the foregoing words are to us, they probably knew a lot more about hyenas, cutting with stones and the best parts of a carcass than most of us would]. In evolving the cognitive ability to understand increasingly detailed relationships, hominids eventually created the means to tag the nouns and actions with arbitrary symbols as a means of information compression for a variety of purposes: initially as internal (neural) words to aid in DMR construction and cortical prediction, but presumably later as uttered words to communicate. Later still, we evolved the means to flexibly combine symbols in a meaningful way, which requires a higher order of intelligence (symbol manipulation) than would (perhaps) seem to be present in most other animals. It appears that Chaser the border collie and Alex the parrot learned arbitrary symbols (human words) and were able to manipulate learned relationships amongst these symbols in linguistic-like ways, but the neuronal mechanisms underlying their learned associations, might bear little resemblance to the

neocortical mechanisms that today support human language and our ability to flexibly manipulate and recombine an endless variety of internal symbols.

From Object Recognition to Language. The learning of verbs and nouns has to be overlaid on the Object Recognition / Invariant Representation systems, working with association cortex to link motivated signs (e.g. the gasp of touching a very hot rock near the fire) and the neural words that represent them and thus define objects, categories and actions. A lot of grammar is spatial in nature, starting with the 3D and 4D physical worlds (lots of prepositions), which then leads to the construction of more complex relationships. One can imagine that a lot of the elements that one might use in a "universal grammar" are already present in a "universal cognitive skill set", via e.g. the learned linking of different elements of our stored experiences: actions, objects, relationships, as e.g. dominance hierarchies, members of a pride or pack, time (hourly, daily, seasonal), spatial relationships (up, down, compass direction, patchy locations of food), causality, serial order as e.g. in construction (nests, beaver dams, sleeping trees). At the dawn of language, protogrammar is the implementation of different kinds of relationships. Stones can be thrown or chipped; fires can go out, objects can be collected that include: rocks, shafts, vines, fur, bones, etc. We do not know the language capabilities of Australopithecus or the earliest homos, but it would be a mistake to suggest that they were not at least as capable as modern chimps, Alex the parrot, or Chaser the border collie. The reason none of these other species evolved an extensive, syntactic language is most uncertain, but one reason that today, Alex, Chaser and other animals can acquire large vocabularies is that they all had TEACHERS. The only animal cultures today with teachers and other specialists are humans. Mammalian neocortex (and avian forebrain) are capable of acquiring and using extended, combinatorial vocabularies, but only humans do so. Animals can solve many problems, as in the case of the amazing crow who solved a 3 step problem to get an extraction tool (Nova or Nature, 2010). Rats and monkeys solve innumerable problems in labs. Language is an extension of all of these more ancestral skills.

Eventually, as our culture progressed, it became valuable to convey more sophisticated kinds and abstractions of information and this required the ability to convey information via word sequence, as in "ball dropped" vs. "dropped ball", with the latter specifically implying (in English) that some agent deliberately or unintentionally dropped a ball (and with possible further implications if it concerned a person's management skills); the first sequence only implies a physical event, even and this distinction is clear even to an 8-year old child. But this is a modern interpretation based on the vast richness of current human language which involves a rat's nest of categorizations (many-to-one) and expansions (one-to-many) (Carpenter and Grossberg, 2008) interconnected in myriad ways—not to mention entire worlds of non-linguistic experiences (memories) associated with EACH of the words we use. Try replaying every memory of yours that involves the word "dog"—ranging from your first pet, to discussions of closing the local dog (greyhound) track (a Boston thing) to the musical question of Who let the Dogs out? This would take you awhile and this modern problem is a serious 10k-connectivity problem! So, in thinking about our ability to convey information by the structure of word sequences (which seems a different problem than being able to merely sequence things) we have to consider the extent to which our recollecting of experiences and cognitively relating and formulating them in terms of either understanding past events or planning future courses of actions led down the road to the evolution of grammar. Alternatively, it may be that our cognitive experiences themselves were enhanced by an ability to symbolize things and then operate on these compressed representations (Neural Words?) in ways that we could not operate on if we had to e.g. manipulate our entire memory database associated with dog or car (or fire or large mammal or stone). To try and think more clearly about this, I think it is better to try and understand the early hominid world and lifestyle, and to try and understand the earliest steps of word creation (in neuronal circuit terms) along with simple word sequencing and how that might convey information—keeping in mind that the earliest hominids

STARTED OUT with most or all of the general intelligence capabilities of all other animals today (but not specialized skills like echolocation or sniffing other dogs butts).

Symbolic Neuronal Information Processing (SNOPS).
Our hypothesized Neural Words are compact codes – in some sense (still to be defined) they are more compact than our invariant representations, with all of their associated neuronal networks. The Neural Words, because of their compactness, are rather more "symbolic" than the larger object representations, in the same sense that a radiation warning sign outside a facility is more compact and symbolic than the nuclear power plant that it represents. Compactness would seem to be useful for both cortical prediction and for the storage of cortical experiences in our Daily Memory Records. Given the similarities of neocortex across mammals, and especially across primates, it seems likely that primates have some level of cortical prediction and daily memory skills that are at least related to what humans have (although this should be tested). But they do not have works of Shakespeare. The Neural Word is thus viewed as step towards linguistic words and an increasing level of abstractness and symbolism. **Language** is thus a construct that may constitute the true pinnacle of human intelligence, opening up an abstract domain of thought unrivaled in the animal kingdom. But an alternative view is that language is primarily a kind of neuronal-shorthand, a set of symbolic tags that reflects a majestically impressive set of non-linguistic neuronal operations whose results emerge into our thoughts and mind via our neuronal consciousness-generating mechanisms. So, are modern grammars and linguistic thought the machinery of our intelligence or just tags of underlying operations? Linguistic tags and language itself are both generated by subconscious information processing machinery, as our neuronal computations roll along moment by moment, but this does not mean that linguistic words and other symbols do not operate in their own architectural (neuronal) domains conferring capabilities never seen before. This is a crucial issue and one that will be explored in the next sections of this book. But in either case, the burning question for me remains this: why were we hominids the ONLY animals to have evolved an extensive, combinatorial symbolic code? To address this we must examine the hominid lifestyle and consider possible new developments in neocortical architectures.

Chapter Summary and a Possible Path to Consciousness
Because of our hyper-associative memory, the brain's cortical prediction problem is like that of playing 1000 chess games at once and needing to forecast or play out every potential move to 10 steps ahead —and with an additional wrinkle of all 1000 chess games being interconnected. This is going to impact every aspect of brain function, from learning higher-order stats (Adams; www.syndar.org), to planning, to storing memories, and to retrieving specific memories (from amongst a vast set of them) and particularly the specific memories that are most important at any given moment. But somehow the brain solves this problem and optimizes our performance in a virtually endless variety of scenarios. This is pretty intelligent. The job of master orchestrator (decision maker) is often assigned to prefrontal cortex (PFC, **Figure 1**), but the basal ganglia also play a crucial role in initiating movements and perhaps thoughts, so the basal ganglia may be the brain's ultimate WTA device: a super-decider that chooses which threads to follow within the chess mindstorm of our brains. This further ties into consciousness, because consciousness can be viewed as a vertebrate decision-making device or process that acts upon the items brought into our stream of consciousness by (e.g.) the Baars and Franklin (2003) consciousness-generating mechanisms. One goal of mine is to reduce Baars and Franklin's *Global Workspace Theory* of consciousness down to the neuronal-circuit level and to mesh it with the content in this chapter and elsewhere in this book. Decision making will be addressed in detail in a separate chapter, but bad decisions and judgment can reflect quite negatively upon one's perceived intelligence.

Bibliography

Ayers J, Witting J (2007) Biomimetic approaches to the control of underwater walking machines. Philos. Transact. A Math Phys. Eng. Sci. 365:273-295.

Baars BJ, Franklin S (2003) How conscious experience and working memory interact. Trends Cogn. Sci. 7:166-172.

Baars BJ, Franklin S (2007) An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA. Neural Networks 20:955-961.

Bar M (2009) Predictions: a universal principle in the operation of the brain. Philos. Trans. Royal Soc. B 364:1181-1182.

Baum (2004) What is Thought? MIT Press, Cambridge MA. 478 pp.

Berkowitz A, Roberts A, Soffe SR (2010) Roles for multifunctional and specialized spinal interneurons during motor pattern generation in tadpoles, zebrafish larvae, and turtles. Front. Behav Neurosci. Jun 28;4:36.

Bickerton D (2009) Adam's Tongue: How humans made language, how language made humans. Hill and Wang Publishers, New York, 286 pp.

Bowers JS (2010) More on grandmother cells and the biological implausibility of PDP models of cognition: a reply to Plaut and McClelland (2010) and Quian Quiroga and Kreiman (2010). Psychol. Rev. 117:300-308.
.
Bowers JS (2009) On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. *Psychol. Rev.* **116**:220-251.

Brady, T.F., Konkle, T., Alvarez, G.A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. PNAS 105:14325-14329.

Brodal A (1981) Neurological Anatomy in relation to clinical medicine, 3rd edition. Oxford University Press, USA, 1072pp.

Burling R (2007) The talking ape: how language evolved. Oxford University Press, USA, 298 pp.

Cajal SRY (1995) Histology of the Nervous System of Man and Vertebrates (History of Neuroscience, No 6). Translation by N. Swanson and L. Swanson. Oxford University Press, 1672 pp.

Cardona A, Saalfeld S, Preibisch S, Schmid B, Cheng A, Pulokas J, Tomancak P, Hartenstein V (2010) An integrated micro- and macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy. PLoS Biol. 8(10)pii:e1000502.

Carla L, Bruna F, Lucianna M, Fernanda L, Paolo S, Luciana C (2010) Distribution of glial cell line-derived neurotrophic factor receptor alpha-1 in the brain of adult zebrafish. J. Anat. 217:174-185.

Carpenter GA, Grossberg S (2008) Adaptive Resonance Theory, in *Encyclopedia of Machine Learning*, C. Sammut & G. Webb, Eds., Springer, p. 1-23.

Chauvin Y, Rumelhart DE (1995) Backpropagation: theory, architectures and applications. Lawrence Erlbaum Associates, Hillsdale NJ, 561 pp.

Christiansen MH, Kirby S (2003) Language evolution: consensus and controversies. Trends Cogn. Sci. 7:300-307.

Courbage M, Nekorkin VI, Vdovin LV (2007) Chaotic oscillations in a map-based model of neural activity. Chaos 17:043109.

Dayan P, Abbott LF (2005) Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. MIT Press, Cambridge MA.

Diamond J (1992) The third chimpanzee. Harper Collins Publishers, NY, 407pp.

Dougherty KJ, Kiehn O (2010) Functional organization of V2a-related locomotor circuits in the rodent spinal cord. Ann. NY Acad. Sci. 1198:85-93.

Fiser J, Aslin RN (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. Psychol. Sci. 12:499-504.

Fodor (1975) The language of thought. Harvard University Press, Cambridge MA.

Gioioso V and O'Malley DM (2009) Conscious Record Memory: Thoughts on an associative, flash-memory system that underlies human intelligence. Soc. Neurosci. Abstracts 2009, #92.22.

Golomb D, Rubin N, Sompolinsky H (1990) Willshaw model: Associative memory with sparse coding and low firing rates. Phys. Rev. A. 41:1843-1854.

Gottfredson L (1997) Mainstream science on intelligence: An editorial with 52 signatories, history, and Intelligence 24:13-23.

Grossberg (1988) Neural Networks and Natural Intelligence. MIT Press, Cambridge MA.

Grossberg S, Pearson LR (2008) Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: toward a unified theory of how the cerebral cortex works. Psychol. Rev. 115:677-732.

Grossmann KS, Giraudin A, Britz O, Zhang J, Goulding M (2010) Genetic dissection of rhythmic motor networks in mice. Prog. Brain Res. 187:19-37.

Gütig R, Sompolinsky H (2009) Time-warp invariant neuronal processing. PloS Biology 7:e1000141, 14 pp.

Hawkins J (2004) On Intelligence. Henry Holt and Company, NY, 261 pages.

Hertz JA, Krogh AD, Palmer RG (1991) Introduction to the theory of neuronal computation. Westview Press, Perseus Books Group, 327 pp.

Hubbard EM, Ramachandran VS (2005) Neurocognitive mechanisms of synesthesia.

Neuron 48:509-520.

Hubel DH and Wiesel TN (1959) Receptive fields of single neurones in the cat's striate cortex. J. Physiol. 148:574-591.

Izhikevich EM (2006) Dynamical systems in neuroscience: the geometry of excitability. MIT Press, Cambridge, MA, 457pp.

Kerr JN, Greenberg D, Helmchen F (2005) Imaging input and output of neocortical networks in vivo. Proc. Natl. Acad .Sci. USA 102:14063-14068.

Legg S, Hutter M (2007) A collection of definitions of intelligence. IDSIA-07-07 Proceedings Technical Report pp. 1-12.

Li N, DiCarlo JJ (2010) Unsupervised natural visual experience rapidly reshapes size-invariant object representation in inferior temporal cortex. Neuron 67:1062-1075.

Li Y, Van Hooser SD, Mazurek M, White LE, Fitzpatrick D (2008) Experience with moving visual stimuli drives the early development of cortical direction selectivity. Nature456:952-956.

Liao JC, Fetcho JR (2008) Shared versus specialized glycinergic spinal interneurons in axial motor circuits of larval zebrafish. J. Neurosci. 28:12982-12992.

Lyu S, Farid H (2003) Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines. Lecture Notes in Computer Science 2578:340-354.

MacNeil MA, Masland RH (1998) Extreme diversity among amacrine cells: implications for function. Neuron 20:971-982

Maimon G, Assad JA (2009) Beyond Poisson: increased spike-time regularity across primate parietal cortex. Neuron 62:426-440.

Micheva KD, O'Rourke N, Busse B, Smith SJ (2010) Array tomography: rodent brain fixation and embedding. Cold Spring Harb Protoc. 2010(11):prot5523. EM reconstruct.

Minsky M (1988) The Society of Mind. Simon and Schuster Publisher.

Minsky M, Papert S (1969) Perceptrons. MIT Press, Cambridge MA, 263 pp.

Mullins OJ, Hackett JT, Buchanan JT, Friesen WO (2010) Neuronal control of swimming behavior: Comparison of vertebrate and invertebrate model systems. Prog Neurobiol. 2010 Nov 18. [Epub ahead of print].

Nieuwenhuys R (1998) The Central Nervous System of Vertebrates. Springer, 2219 pp.

O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 34:171–175.

Pepperberg IM (2002) The Alex Studies: Cognitive and Communicative Abilities of Grey Parrot. Harvard University Press, Cambridge MA, 434 pp.

Pepperberg IM (2009) Alex & Me: How a Scientist and a Parrot Discovered a Hidden World of Animal Intelligence--and Formed a Deep Bond in the Process. Harper Collins Publisher. 288 pp.

Pinker S (1999) How the mind works. Ann. NY Acad. Sci. Wiley Press.

Plaut DC, McClelland JL (2010) Locating object knowledge in the brain: comment on Bowers's (2009) attempt to revive the grandmother cell hypothesis. Psychol. Rev. 117:284-288.

Pulvermüller F, Knoblauch A (2009) Discrete combinatorial circuits emerging in neural networks: a mechanism for rules of grammar in the human brain?   Neural Netw. 22:161-172.

Purves D, et al. (2007) Neuroscience, 4th Edition. Sinauer Associates, Sunderland MA, 857 pp.

Quian Quiroga R, Kreiman G (2010) Measuring sparseness in the brain: comment on Bowers (2009). Psychol. Rev. 117:291-297.

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. Nature. 2005 Jun 23;435(7045):1102-7.

Quiroga RQ, Kreiman G, Koch C, Fried I (2008) Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends Cogn. Sci.* **12**:87-91.

Rendall D, Owren MJ, Ryan MJ (2009) What do animal signals mean?
Animal Behaviour 78:233-240.   [notes that Shannon IT fails in animal communications]

Rieke F, Warland D, de Ruyter von Steveninck R, Bialek W (1999) Spikes: Exploring the Neural Code. MIT Press, Cambridge, MA.

Rizzolatti G, Craighero L (2004) The mirror-neuron system. Ann. Rev. Neurosci. 27:169–192.

Rust NC, DiCarlo JJ (2010) Selectivity and tolerance ("invariance") both increase as visual information propagates from cortical area V4 to IT. J. Neurosci. 30:12978-12995.

Seung HS (2009) Reading the book of memory: sparse sampling versus dense mapping of connectomes. Neuron 62:17-29.

Shi B, Bourne J, Harris KM (2011) SynapticDB, Effective Web-based Management and Sharing of Data from Serial Section Electron Microscopy. Neuroinformatics, *in press.*

Singer T, McConnell MJ, Marchetto MC, Coufal NG, Gage FH (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? Trends Neurosci. 33:345-354.

Stevens CF (1998) Neuronal diversity: too many cell types for comfort? Curr. Biol. 8:R708-R710.

Sugino K, Hempel CM, Miller MN, Hattox AM, Shapiro P, Wu C, Huang ZJ, Nelson SB (2006) Molecular taxonomy of major neuronal classes in the adult mouse forebrain. Nat. Neurosci. 9:99-107.

Talpalar AE, Kiehn O (2010) Glutamatergic mechanisms for speed control and network operation in the rodent locomotor CPG. Front. Neural Circuits Vol. 4, Article 19, 14pp.

Tseng GF, Prince DA (1993) Heterogeneity of rat corticospinal neurons. J. Comp. Neurol. 335:92-108.

Weir AAS, Chappell J and Kacelnik A (2002) Shaping of hooks in New Caledonian crows. Science. 297:981.

Xie X, Seung HS (2003) Equivalence of backpropagation and contrastive Hebbian learning in a layered network. Neural Comput. 15:441-454.

Zelenin PV (2011) Reticulospinal neurons controlling forward and backward swimming in the lamprey. J. Neurophysiol. Jan 19. [Epub ahead of print].

These next two items will be added to Bibliography
They were left in draft form partly for time reasons, but also to note the number of times they are cited (almost 4000 times each), which is astronomical. Their impact is thus enormous, and I am guessing they did OK in sales.

[BOOK] The language of thought
JA Fodor - 1975 - books.google.com
Jerrold J. Katz D. Terence Langendoen George A. Miller SERIES EDITORS **...** THE LANGUAGE OF THOUGHT)ERRyA.PODOR Harvard University Press Cambridge, Massachusetts **...** The Language & Thought Series SERIES EDITORS JERROLD J KATZ. THE GRADUATE **...**
Cited by 3922 - Related articles - Library Search - All 4 versions

How the mind works
S Pinker - Annals of the New York Academy of Sciences, 1999 - Wiley Online Library
The human mind is a remarkable organ. It has allowed us to walk on the moon, to discover the physical basis of life and the universe, and to play chess almost as well as a computer. But the brain raises a paradox. On the one hand, many tasks that we take for granted-walking **...**
Cited by 3797 - Related articles - Library Search - BL Direct - All 21 versions

Addendums (Post-Sinauer Submission)

1. Excerpt below on Wolf Singer describes contributions of neural oscillators to DMRs; did not get to this in the sample chapter, and **how neural oscillators are central to Object Recognition!** [which should be integrated into this chapter, space permitting].

2. Adam Reeves mentioned the **great speed of object recognition.**
--cannot be faster because of my theory that "less is going on" in narrow experimental task because everything is parallel—its not like when DOD says its 9 woman months to have a baby; hire 9 women and do it in a month.
-- but maybe we have an iconic encyclopedia of objects and we can match them VERY fast, with dot stereograms or whatever
-- did Quiroga report on how fast Jennifer Aniston lit up their neurons?

3. Additional Supplements: see file NLM.Appendix.02 for extensive materials on corvids, Wolf Singer, oscillations, open questions and more (in extended version not available online).