Where artificial intelligence and neuroscience meet: The search for grounded architectures of cognition

Velde, F. van der

Hindawi Publishing Corporation
Advances in Artificial Intelligence
Volume 2010, Article ID 918062, 18 pages
doi:10.1155/2010/918062

Review Article

Where Artificial Intelligence and Neuroscience Meet: The Search for Grounded Architectures of Cognition

Frank van der Velde

Cognitive Psychology, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands

Correspondence should be addressed to Frank van der Velde, vdvelde@fsw.leidenuniv.nl

Received 31 August 2009; Revised 11 November 2009; Accepted 12 December 2009

Academic Editor: Daniel Berrar

The collaboration between artificial intelligence and neuroscience can produce an understanding of the mechanisms in the brain that generate human cognition. This article reviews multidisciplinary research lines that could achieve this understanding. Artificial intelligence has an important role to play in research, because artificial intelligence focuses on the mechanisms that generate intelligence and cognition. Artificial intelligence can also benefit from studying the neural mechanisms of cognition, because this research can reveal important information about the nature of intelligence and cognition itself. I will illustrate this aspect by discussing the grounded nature of human cognition. Human cognition is perhaps unique because it combines grounded representations with computational productivity. I will illustrate that this combination requires specific neural architectures. Investigating and simulating these architectures can reveal how they are instantiated in the brain. The way these architectures implement cognitive processes could also provide answers to fundamental problems facing the study of cognition.

1. Introduction

Intelligence has been a topic of investigation for many centuries, dating back to the ancient Greek philosophers. But it is fair to say that it is a topic of a more scientific approach for just about 60 years. Crucial in this respect is the emergence of artificial intelligence (AI) in the mid 20th century. As the word "artificial" suggests, AI aimed and aims not only to understand intelligence but also to build intelligent devices. The latter aim adds something to the study of intelligence that was missing until then: a focus on the mechanisms that generate intelligence and cognition (here, I will make no distinction between these two concepts).

The focus on mechanisms touches upon the core of what intelligence and cognition are all about. Intelligence and cognition are about mechanisms. Only a true mechanistic process can transform a sensory impression into a motor action. Without it, cognition and intelligence would not have any survival value. This is quite clear for processes like pattern recognition or motor planning, but it also holds for "higher" forms of intelligence (cognition), like communication or planning. Consequently, a theory of a cognitive process that does not describe a true mechanism (one that, at least in principle, can be executed) is not a full theory of that process, but at best an introduction to a theory or a philosophical account.

In this respect, AI is not different from other sciences like physics, chemistry, astronomy, and genetics. Each of these sciences became successful because (and often when) they focussed on an understanding of the mechanisms underlying the phenomena and processes they study. Yet, the focus on mechanisms was not always shared by other sciences that study intelligence or cognition, like psychology or neuroscience. For the most part, psychology concerned (and still concerns) itself with a description of the behavior related to a particular cognitive process. Neuroscience, of course, studied and studies the physiology of neurons, which aims for a mechanistic understanding. Yet, for a long time it stopped short at a translation from physiology to cognition.

However, the emergence of cognitive neuroscience in the 1990s introduced a focus on a mechanistic account of natural intelligence within neuroscience and related sciences. Gazzaniga, one of the founders of cognitive neuroscience, makes this point explicitly: "At some point in the future,

cognitive neuroscience will be able to describe the algorithms that drive structural neural elements into the physiological activity that results in perception, cognition, and perhaps even consciousness. To reach this goal, the field has departed from the more limited aims of neuropsychology and basic neuroscience. Simple descriptions of clinical disorders are a beginning, as is understanding basic mechanisms of neural action. The future of the field, however, is in working toward a science that truly relates brain and cognition in a mechanistic way." [1, page xiii].

It is not difficult to see the relation with the aims of AI in this quote. Gazzaniga even explicitly refers to the description of "algorithms" as the basis for understanding how the brain produces cognition. Based on its close ties with computer science, AI has always described the mechanisms of intelligence in terms of algorithms. Here, I will discuss what the algorithms as intended by Gazzaniga and the algorithms aimed for by AI could have in common. I will argue that much can be gained by a close collaboration in developing these algorithms. In fact, a collaboration between cognitive neuroscience and AI may be necessary to understand human intelligence and cognition in full.

Before discussing this in more detail, I will first discuss why AI would be needed at all to study human cognition. After all, (cognitive) neuroscience studies the (human) brain, and so it could very well achieve this aim on its own. Clearly, (cognitive) neuroscience is crucial in this respect, but the difference between human and animal cognition does suggest that AI has a role to play as well (in combination with (cognitive) neuroscience. The next section discusses this point in more detail.

2. Animal versus Human Cognition

Many of the features of human cognition can be found in animals as well. These include perception, motor behavior and memory. But there are also substantial differences between human and animal cognition. Animals, primates included, do not engage in science (such as neuroscience or AI) or philosophy. These are unique human inventions. So are space travel, telescopes, universities, computers, the internet, football, fine cooking, piano playing, money, stock markets and the credit crisis, to name but a few.

And yet, we do these things with a brain that has many features in common with animal brains, in particular that of mammals. These similarities are even more striking in case of the neocortex, which is in particular involved in cognitive processing. In an extensive study of the cortex of the mouse, Braitenberg [2] and Braitenberg and Schüz [3] observed striking similarities between the cortex of the mouse and that of humans. In the words of Braitenberg [2, page 82]: "All the essential features of the cerebral cortex which impress us in the human neuroanatomy can be found in the mouse too, except of course for a difference in size by a factor 1000. It is a task requiring some experience to tell a histological section of the mouse cortex from a human one. . . . With electronmicrographs the task would actually be almost impossible."

It is hazardous to directly relate brain size to cognitive abilities. But the size of the neocortex is a different matter. There seems to be a direct relation between the size of the neocortex and cognitive abilities [4]. For example, the size of the human cortex is about four times that of chimpanzees, our closest relatives. This difference is not comparable to the difference in body size or weight between humans and chimpanzees.

So, somehow the unique features of human cognition are related to the features of the human cortex. How do we study this relation? Invasive animal studies have been extremely useful for understanding features of cognition shared by animals and humans. An example is visual perception. Animal research has provided a detailed account of the visual cortex as found in primates (e.g., macaques [5]). Based on that research, AI models of perception have emerged that excel in comparison to previous models [6]. Furthermore, neuroimaging research begins to relate the structure of the visual cortex as found in animals to that of humans [7].

So, in the case of visual perception we have the ideal combination of neuroscience and AI, producing a mechanistic account of perception. But what about the unique features of human cognition?

In invasive animal studies, electrodes can penetrate the cortex at arbitrary locations, the cortex can be lesioned at arbitrary locations, and the animal can be sacrificed to see the effects of these invasions. On occasion, electrodes can be used to study the human cortex, when it is invaded for medical reasons [8]. But the rigorous methods as used with animals are not available with humans. We can use neuroimaging, but the methods of neuroimaging are crude compared to the methods of animal research. EEG (electroencephalogram) provides good temporal resolution but its spatial resolution is poor. For fMRI (functional magnetic resonance imaging), the reverse holds. So, these methods on their own will not provide us with the detailed information provided by animal research.

This is in particular a problem for studying the parts of the human brain that produce space travel, telescopes, universities, computers, the internet, football, fine cooking, piano playing, money, stock markets and the credit crisis, if indeed there are such distinguishable parts. It is certainly a problem for studying the parts of the human brain that produce language and reasoning, which are at the basis of these unique human inventions. For these aspects of cognition, there is no animal model that we can use as a basis, as in the case of visual perception. (Indeed, if there were such animal models, that is, if animal cognition was on a par with human cognition, we would have to question the ethical foundations of doing this kind of research.)

So, not surprisingly, our knowledge of the neural mechanisms of language or reasoning is not comparable to that of visual perception. In fact, we do not have neural models that can account for even the basic aspects of language processing or reasoning.

In his book on the foundation of language, Jackendoff [9] summarized the most important problems, the "four challenges for cognitive neuroscience", that arise with a neural implementation of combinatorial structures, as found in

human cognition. These challenges illustrate the difficulties that occur when combinatorial hierarchical structures are implemented with neural structures. Consider the first two challenges analyzed by Jackendoff.

The first challenge concerns the massiveness of the binding problem as it occurs in language, for example in hierarchical sentence structures. For example, in the sentence *The little star is besides the big star*, there are bindings between adjectives and nouns (e.g, *little star* versus *big star*), but also bindings between the noun phrase *the little star* and the verb phrase *is besides the big star* or between the prepositional phrase *besides the big star* and verb *is*.

The second challenge concerns the problem of multiple instantiations, or the "problem of 2", that arises when the same neural structure occurs more than once in a combinatorial structure. For example, in the sentence *The little star is besides the big star*, the word *star* occurs twice, first as subject of the sentence and later as the noun of the prepositional phrase.

These challenges (and the other two) were not met by any neural model at the time of Jackendoff's book. For example, consider synfire chains [10]. A synfire chain can arise in a feedforward network when activity in one layer cascades to another layer in a synchronous manner. In a way, it is a neural assembly, as proposed by Hebb [11] with a temporal dimension added to it [3]. Synfire chains have sometimes been related to compositional processing [12], which is needed in the case of language.

But is clear that synfire chains do not meet the challenges discussed by Jackendoff. For example, in *The little star is besides the big star* a binding (or compositional representation) is needed for *little star* and *big star*, but not for *little big star* (this noun phrase is not a part of the sentence). With synfire chains (and Hebbian assemblies in general [13]), we would have synfire chains for *star*, *little* and *big*. The phrase *little star* would then consist of a binding (link) between the synfire chains for *little* and *star*. At the same time, the phrase *big star* would consist of a binding between the synfire chains for *big* and *star*. However, the combination of the bindings between the synfire chains for *little*, *big* and *star* would represent the phrase *little big star*, contrary to the structure of the sentence.

This example shows that synfire chains fail to account for the "problem of two". Because the word *star* occurs twice in the sentence, somehow these occurrences have to be distinguished. Yet, a neural representation of a concept or word, like *star*, is always the same representation (in this case the same synfire). Indeed, this is one of the important features of neural cognition, as I will argue below. But this form of conceptual representation precludes the use of direct links between synfire chains (or assemblies) as the basis for the compositional structures found in language (see [13] for a more extensive analysis).

3. Investigating the Neural Basis of Human Cognition

Given the additional difficulties involved in studying the neural basis of the specific human forms of cognition, as outlined above, the question arises how we can study the neural basis of human cognition.

Perhaps we should first study the basic aspects of neural processing, before we could even address this question. That is, the study of human forms of cognition would have to wait until we acquire more insight into the behavior of neurons and synapses, and smaller neural circuits and networks.

However, this bottom-up approach may not be the most fruitful one. First, because it confuses the nature of understanding with the way to achieve understanding. In the end, a story about the neural basis of human cognition would begin with neurons and synapses (or even genes) and would show how these components form neural circuits and networks, and how these structures produce complex forms of cognition. This is indeed the aim of understanding the neural basis of human cognition. But is not necessarily the description of the sequence in which this understanding should or even could be obtained.

A good example of this difference is found in the study of the material world. In the end, this story would begin with an understanding of elementary particles, how these particles combine to make atoms, how atoms combine to make molecules, how molecules combine to make fluids, gases and minerals, how these combine to make planets, how planets and stars combine to make solar systems, how these combine to make galaxies, and how galaxies combine to form the structure of the universe.

This may be the final aim of understanding the material world, but it is not the way in which this understanding is achieved. Physics and astronomy did not begin with elementary particles, or even atoms. In fact, they began with the study of the solar system. This study provided the first laws of physics (e.g., dynamics) which could then be used to study other aspects of the material world as well, such as the behavior of atoms and molecules. The lesson here is that new levels or organization produce new regularities of behavior, and these regularities can also provide information about the lower levels of organization. Understanding does not necessarily proceed from bottom to top, it can also proceed from top to bottom.

Perhaps the best way to achieve understanding is to combine bottom-up and top-down information. The discussion above about the foundations of language provides an example. We can study language (as we can study planets) and obtain valuable information about the structure of language. This information then sets the boundary conditions, such as the two challenges discussed above, that need to be fulfilled in a neural account of language structure. In fact, these boundary conditions provide information that may be difficult to come by in a pure bottom-up approach.

The study of the material world also provides information of how the interaction between the bottom-up and to-down approach might proceed. Astronomy studies objects (stars and galaxies) that are in a way inaccessible. That is we cannot visit them or study them in a laboratory setting. In a way, this resembles the study of the human brain, which is inaccessible in the sense that we cannot do the rigorous experiments as we do with animals.

Yet, astronomy has acquired a profound understanding of stars and galaxies. It can, for example, describe the evolution of stars even though that proceeds over millions of years. In the 19th century, however, astronomy was still restricted to describing the position of stars and their relative magnitude. But physics can study the properties of matter in a laboratory. Combined with theoretical understanding (e.g., quantum physics), it can show how light provides information about the structure of matter. This information can be used to study the properties of stars as well. Furthermore, theoretical understanding of matter (e.g., statistical physics) can also provide information about how stars could evolve, which in turn can be investigated with astronomical observations.

In short, the success of astronomy depends on a combination of studying the basics of matter (physics), observing the properties of stars (astronomy) and combining these levels with theoretical analysis. In this three-fold combination, each component depends on the other. As a result, seemingly inaccessible phenomena can be studied and understood on a substantial level of complexity.

A similar approach could be successful in studying the seemingly inaccessible neural basis of human cognition (as exemplified in language and reasoning). That is, detailed investigation of basic neural structures, observations of brain processes based on neuroimaging, and theoretical or computational research which investigates how cognitive processes as found in humans can be produced with neural structures and how the behavior of these structures can be related to observations based on neuroimaging. As in the case of astronomy, each of these components is necessary. But the role of AI will be restricted to the computational part. So, I will focus on that in the remainder of this paper.

4. Large-Scale Simulations

An important development in the collaboration between AI and neuroscience is the possibility of large-scale simulations of neural processes that generate intelligence. For example, the mouse cortex has approximately 8×10^6 neurons and 8000 synapses per neuron. Recently, an IBM research group represented 8×10^6 neurons and 6400 synapses per neuron on the IBM Blue Gene processor, and ran 1 s of model time in 10 s of real time [14]. With this kind of computing power, and its expected increase over the coming years, it can be expected that large sections of the human cortex (which is about 1000 times larger than the mouse cortex [3]) can be modelled in comparable detail in the near future.

These large-scale simulations will provide a virtual research tool by which characteristics of the human brain, and their relation to cognitive function, can be investigated on a scale and level of detail that is not hampered by the practical and ethical limitations of (invasive) brain research. For example, large-scale simulations can be used to study the interaction between thousands of neurons in realistic detail, or to investigate the effect of specific lesions on these interactions, or to investigate the role of specific neurotransmitters on neuronal interactions. In this way, the limitations of experimental methods can be augmented. No experimental method gives detailed information about the interaction of thousands of neurons, and no experimental method can vary parameters in the interaction at will to study their effect. The Blue Brain Project [15] is an attempt to study how the brain functions in this way, and to serve as a tool for neuroscientists and medical researchers.

But the Blue Brain Project is focused on creating a physiological simulation for biomedical applications. By its own admission, it is not (yet) an artificial intelligence project. However, from an AI perspective, large-scale simulations of neural processes can be used as a virtual laboratory to study the neural architectures that generate natural intelligence and cognition. These architectures depend on the structure of the brain, and the neocortex in particular, as outlined below.

4.1. Structure of the Neocortex. In the last decades, a wealth of knowledge has been acquired about the structure of the cortex (e.g., [16]). A comparison of the structure of the cortex in different mammals shows that the basic structure of the cortex in all mammals is remarkably uniform. The one factor that distinguishes the cortex of different mammals is their size. For example, the cortex of humans is about 1000 times that of a mouse, but at a detailed (microscopically) level it is very hard to distinguish the two [3]. This finding suggests that the unique features of human cognition might derive from the fact that more information can be processed, stored and interrelated in the extended networks and systems of networks as found in the human neocortex.

Furthermore, the basic structure of the cortex itself is highly regular. Everywhere within the cortex, neurons are organized in horizontal layers (i.e., parallel to the cortical surface) and in small vertical columns. The basic layered structure consists of six layers, which are organized in three groups: a middle layer (layer 4), the superficial layers (layers above layer 4) and the deep layers (layers below layer 4). The distribution of different kinds of neurons within the layers and columns is similar in all parts of the cortex. More than 70% of all neurons in the cortex are pyramidal neurons. Pyramidal neurons are excitatory, and they are the only neurons that form long-range connections in the cortex (i.e., outside their local environment). The probability that any two pyramidal neurons have more than two synaptic contacts with each other is small. Yet, substantially more than two synaptic inputs are needed to fire a pyramidal neuron. This indicates that neurons in the cortex operate in groups or populations. Furthermore, neurons within a given column in the cortex often have similar response characteristics, which also indicates that they operate as a group or population. In all parts of the cortex, similar basic cortical circuits are found. These circuits consist of interacting populations of neurons, which can be located in different layers.

At the highest level of organization, the cortex consists of different areas and connection structures ("pathways") in which these areas interact. Many pathways in the cortex are organized as a chain or hierarchy of cortical areas. Processing in these pathways initially proceeds in a feedforward manner, in which the lower areas in the hierarchy process input information first, and then transmit it to higher areas in

the hierarchy. However, almost all feedforward connections in the pathways of the cortex are matched by feedback connections, which initiate feedback processing in these pathways. The connection patterns in the pathways, consisting of feedforward, feedback and lateral connections, begin and terminate in specific layers. For example, feedforward connections terminate in layer 4, whereas feedback connections do not terminate in this layer.

An example of the relation between cortical structures and cognitive processing is given by visual perception. Processing visual information is a dominant form of processing in the brain. About 40% of the human cortex is devoted to it (in primates even more than 50%). The seemingly effortless ability to recognize shapes and colors, and to navigate in a complex environment is the result of a substantial effort on the part of the brain (cortex). The basic features of the visual system are known (e.g., [5]). The visual cortex consists of some 30 cortical areas, that are organized in different pathways. The different pathways process different forms of visual information, or "visual features", like shape, color, motion, or position in visual space.

All pathways originate from the primary visual cortex, which is the first area of the cortex to receive retinal information. Information is transmitted from the retina in a retinotopic (topographic) manner to the primary visual cortex. Each pathway consists of a chain or hierarchy of cortical areas, in which information is initially processed in a feedforward direction. The lower areas in each pathway represent visual information in a retinotopic manner. From the lower areas onwards, the pathways begin to diverge.

Object recognition (shape, color) in the visual cortex begins in the primary visual cortex, located in the occipital lobe. Processing then proceeds in a pathway that consists of a sequence of visual areas, going from the primary visual cortex to the temporal cortex. The pathway operates initially as a feedforward network (familiar objects are recognized fast, to the extent that there is little time for extensive feedforward-feedback interaction). Objects (shapes) can be recognized irrespective of their location in the visual field (i.e., relative to the point of fixation), and irrespective of their size.

Processing information about the spatial position of an object occurs in a number of pathways, depending on the output information produced in each pathway. For example, a specific pathway processes position information in eye-centered coordinates, to steer eye movements. Other pathways exist for processing position information in body-, head-, arm- or finger-centered coordinates. Each of these pathways consist of a sequence of visual areas, going from the primary visual cortex to the parietal cortex (and to the prefrontal cortex in the case of eye movements).

5. From Neural Mechanisms to Cognitive Architectures

Although several levels of organization can be distinguished in the brain, ranging from the cell level to systems of interacting neural networks, the neural mechanisms that fully account for the generation of cognition emerge at the level of neural networks and systems (or architectures) of these networks. A number of important issues can be distinguished here.

The structure of the cortex seems to suggest that the implementation of cognitive processes in the brain occurs with networks and systems of networks based on the uniform local structures (layers, columns, basic local circuits) as building blocks. The organization at the level of networks and systems of networks can be described as "architectures" that determine how specific cognitive processes are implemented, or indeed what these cognitive processes are.

Large-scale simulations of these architectures provide a unique way to investigate how specific architectures produce specific cognitive processes. In the simulation, the specific features of an architecture can be manipulated, to understand how they affect the cognitive process at hand. Furthermore, human cognition is characterized by certain unique features that are not found in animal cognition, or in a reduced form only (e.g., as in language, reasoning, planning). These features have to be accounted for in the analysis of the neural architectures that implement human cognitive processes. An interesting characteristic of these architectures is that they would consist of the same kind of building blocks and cortical structures as found in all mammalian brains. Investigating the computational features of these building blocks provides important information for understanding these architectures.

Because the cortex consists of arrays of columns, containing microcircuits, the understanding of local cortical circuits is a prerequisite for understanding the global stability of a highly recurrent and excitatory network as the cortex. An important issue here is whether the computational characteristics of these microcircuits can be characterized by a relatively small number of parameters [17]. A small number of parameters which are essential for the function of local circuits, as opposed to the large number of neural and network parameters, would significantly reduce the burden of simulating large numbers of these circuits, as required for the large-scale simulation of cognitive processes. It would also emphasize the uniform nature of columns as building blocks of the cortex.

Another important issue concerns the computational characteristics of the interaction between feedforward and feedback networks in the cortex. Connections in the feedforward direction originate for the most part in the superficial layers and sometimes in the deep layers, and they terminate in the middle layer (layer 4) of the next area. Within that area, the transformation from input activity (layer 4) to output activity (superficial or deep layers) occurs in the local cortical circuits (as found in the columns) that connect the neural populations in the different layers. Feedback processing starts in the higher areas in a hierarchy and proceeds to the lower areas. Feedback connections originate and terminate in the superficial and deep layers of the cortex.

So, it seems that feedforward activity carries information derived from the outside world (bottom up information), whereas feedback activity is more related to expectations generated at higher areas within an architecture (top-down

expectations). The difference between the role of feedforward activation and that of feedback activation is emphasized by the fact that they initially activate different layers in the cortex. In particular, feedback activation terminates in the layers that also produce the input for feedforward activity in the next area. This suggest that feedback activity (top-down expectation) modulates the bottom-up information as carried by feedforward activity. It is clear that this modulation occurs in the microcircuits (columns) that interconnect the different layers of the cortex, which again emphasizes the role of these circuits and illustrates the interrelation between the different computational features of the cortex.

The large-scale simulation of cortical mechanisms works very well when there is a match between the knowledge of a cortical architecture and the cognitive processes it generates, as in the case of the visual cortex. For example, the object recognition model of Serre et al. is based on cortex-like mechanisms [6]. It shows good performance, which illustrates the usefulness of cortical mechanisms for AI purposes. Also, the model is based on neural networks which could be implemented in parallel hardware, which would increase their processing speed. Moreover, the weight and energy consumption of devices based on direct parallel implementation of networks would be less than that of standard computers, which enhances the usefulness of these models in mobile systems.

So, when a cortical architecture of a cognitive process is (relatively) well known, as in the visual cortex, one could say that AI follows the lead of (cognitive) neuroscience. But not all cortical architectures of cognition are as well known as the visual cortex. Knowledge of the visual cortex derives to a large extent from detailed animal experiments. Because these experiments are not available for cognitive processes that are more typically human, such as language and reasoning, detailed information about their cortical mechanisms is missing.

Given the uniform structure of the cortex, we can make the assumption that the cortical architectures for these cognitive processes are based on the cortical building blocks as described above. But additional information is needed to unravel these cortical architectures. It can be found in the nature of the cognitive processes they implement. Because specific neural architectures in the cortex implement specific cognitive processes, the characteristics of these processes provide information about their underlying neural mechanisms. In particular, the specific features of human cognition have to be accounted for in the analysis and modelling of the neural architectures involved. Therefore, the analysis of these features provides important information about the neural architectures instantiated in the brain.

6. From Cognitive Architectures to Neural Mechanisms

AI might take the lead in the analysis of mechanisms that can generate features of human cognition. So, AI could provide important information about the neural architectures instantiated in the brain when the mechanisms it provides are combined with knowledge of cortical mechanisms. A number of features of (human) cognition can be distinguished where insight in cognitive mechanisms is important to understand the cortical architectures involved.

6.1. Parallel versus Sequential Processing. A cognitive neural architecture can be characterized by the way it processes information. A main division is that between parallel processing of spatially ordered information and processing of sequentially ordered information.

Parallel processing of spatially ordered information is found in visual perception. An important topic in this respect is the location and size invariant identification of objects in parallel distributed networks. How this can be achieved in a feedforward network is not yet fully understood, even though important progress has been made for object recognition (e.g., [6]). An understanding of this ability is important, because visual processing is a part of many cognitive tasks. However, understanding the computational mechanisms of location and size invariant processing in the brain is also important in its own right, given the applications that could follow from this understanding.

Sequentially ordered information is found in almost all forms of cognitive processing. In visual perception, for example, a fixation of the eyes lasts for about 200 ms. Then a new fixation occurs, which brings another part of the environment in the focal field of vision. In this way, the environment is explored in a sequence of fixations. Other forms of sequential processing occur in auditory perception and language processing. Motor behavior also has clear sequential features. The way in which sequentially ordered information can be represented, processed and produced in neural architectures is just beginning to be understood [13]. Given its importance for understanding neurocognition, this is an important topic for further research.

6.2. Representation. Many forms of representation in the brain are determined by a frame of reference. On the input side, the frame of reference is based on the sensory modality involved. For example, the initial frame of reference in visual perception is retinotopic. That is, in the early (or lower) areas of the visual cortex, information is represented topographically, in relation with the stimulation on the retina. On the output side, the frame of reference is determined by the body parts that are involved in the execution of a movement. For example, eye positions and eye movements are represented in eye-centered coordinates. Thus, to move the eyes to a visual target, the location of the target in space has to be represented in eye-centered coordinates. Other examples of (different) "motor representations" are head-, body-, arm-, or finger-centered coordinates. The nature of these representations and the transformations between their frames of reference have to be understood. Three important issues can be distinguished in particular.

The first one concerns the nature of feedforward transformations. When sensory information is used to guide an action, sensory representations are transformed into motor representations. For example, to grasp an object with visual

guidance, the visual information about its location has to be transformed into the motor representations needed to grasp the object. In this case, the transformations to the motor representations start from a retinotopic representation. The question is what the different forms of motor representation are, how the neural transformations between retinotopic representation and these different motor representations proceed, and how they are learned.

The second one concerns the integration of motor systems. An action often involves the movement of different body parts. The question is how these different motion systems are integrated. That is, how are the transformations between different motor representations performed, and how are they learned. In particular, the question is whether the transformations between motor systems are direct (e.g., from head to body representation and vice versa), or whether they proceed through a common intermediary representation. Suggestions have been made that eye-centered coordinates function as such an intermediary representation (lingua franca). In this way, one motor representation is first transformed into eye-centered coordinates before it is transformed into another motor transformation. An answer to this question is also of relevance for visual motor guidance (e.g., the effect of visual attention on action preparation, [18]).

The third one concerns the effect of feedback transformations. These transformations concern the effect of motor planning on sensory (e.g., visual) processing. For example, due to an eye shift a new part of the visual space is projected on a given location of the retina, replacing the previous projection. In physical terms, there is no difference between a new projection on the retina produced by the onset of a new stimulus (i.e., a stimulus not yet present in the visual field), or a new projection on the same retinal location produced by a stimulus (already present in the visual field) due to an eye shift. In both cases, there is an onset of a stimulus on the given retinal location. However, at least some neurons in the visual cortex respond differently to these two situations. The difference is most likely due to the effect of motor planning and motor execution on the visual representation. In case of an eye shift, information is available that a new stimulus will be projected on a given retinal location. This information is absent in the case of a direct stimulus onset (i.e., the onset of a stimulus not yet present in the visual field). Through a feedback transformation, the motor representation related to the eye shift can be transformed into a retinotopic representation, which can influence the representation of the new visual information. The stability of visual space is related to these feedback transformations. Because the body, head and eyes are moving continuously, the retinal projections also fluctuate continuously due to these movements. Yet, the visual space is perceived as stable. Visual stability thus results from an integration of visual and motor information.

6.3. Productivity. A fundamental feature of human cognition is the practically unlimited productivity of human cognition. Cognitive productivity concerns the ability to process or produce a virtually unlimited number of cognitive structures in a given cognitive domain. For example, a virtually unlimited number of novel sentences can (potentially) be understood or produced by a normal language user. Likewise, visual perception provides the ability to navigate in a virtually unlimited number of novel visual scenes (e.g., novel environments like unknown cities).

In the case of visual perception, productivity is found in animals as well. But with language and reasoning, productivity is uniquely human. A conservative estimate shows that humans can understand a set of 10^{20} (meaningful) sentences or more [19, 20]. This kind of productivity is unlimited in any practical sense of the word. For example, the estimated lifetime of the universe is in the order of 10^{17} to 10^{18} seconds. This number excludes that we could learn each sentence in the set of 10^{20}. Instead, we can understand and produce sentences from this set only in a productive manner.

In computational terms, productivity results from the ability to process information in a combinatorial manner. In combinatorial processing, a cognitive structure (e.g., sentence, visual scene) is processed in terms of its components (or constituents) and the relations between the components that determine the overall structure. Sentences are processed in terms of words and grammatical relations. Visual scenes are processed in terms of visual features like shapes, colors, (relative) locations, and the binding relations between these features.

To understand the neural basis of human cognition, it is essential to understand how combinatorial processing is implemented in neural systems as found in the cortex. A recently proposed hypothesis is that all forms of combinatorial processing in neural systems depend on a specific kind of neural architectures [13]. These architectures can be referred to as neural "blackboard" architectures. They consist of specialized networks that interact through a common neural blackboard.

An example is found in the visual cortex. Visual features like shape, color, motion, position in visual space, are processed and identified in specialized (feedforward) networks. Through feedback processing and interaction in the lower retinotopic areas of the visual cortex, these specialized networks can interact. In this way, the (binding) relations between the visual features of an object can be established [18]. The structure of the neural blackboard architecture for vision is determined by the kind of information it processes, in particular the fact that visual information is (initially) spatially ordered. The characteristics of visual (spatial) information thus provide information about the structure of the neural architecture for vision.

In a similar way, the characteristics of sequentially ordered information, for example, as found in language, or reasoning, or motor planning, and so forth, can be used to determine the structure of the neural architectures involved in these forms of processing. Because combinatorial processing imposes fundamental constraints on neural architectures, these constraints can be used to generate hypotheses about the underlying brain structures and dynamics. In particular, when they are combined with the nature of conceptual representation, as discussed in the next section.

7. Grounded Architectures of Cognition

A potential lead of AI in analyzing the mechanisms of cognition is perhaps most prominent with cognitive processes for which no realistic animal model exists. Examples are language, detailed planning and reasoning. A fascinating characteristic of these processes is that they are most likely produced with the same cortical building blocks as described earlier, that is, the cortical building blocks that also produce cognitive processes shared by humans and animals, such as visual perception and motor behavior.

Apparently, the size of the neocortex plays a crucial role here. The human cortex is about four times the size of that of a chimpanzee, 16 times that of a macaque monkey and a 1000 times that of a mouse [3, 4]. Given the similarity of the structure of the cortex, both within the cortex and between cortices of different mammals this relation between size and ability makes sense. Having more of the same basic cortical mechanisms available will make it easier to store more information, but apparently it also provides the ability to recombine information in new ways.

Recombining information is what productivity is about. So, we can expect these more exclusively human forms of cognition to be productive. But the way information is stored should be comparable with the way information is stored in the brain in all forms of cognition. Examples are the forms of representation found in the visual cortex or the motor cortex, as discussed above. This is a challenge for AI and cognitive science: how to combine productivity as found in human cognition with the forms of representation found in the brain. Solving this challenge can provide important information about how these forms of cognition are implemented in the brain. It can also provide information about the unique abilities of human cognition which can be used to enhance the abilities of AI.

To understand the challenge faced by combining cognitive productivity with representation as found in the brain, consider the way productivity is achieved in the classical theory of cognition, or classical cognitivism for short, that arose in the 1960s. Classical cognitive architectures (e.g., [21, 22]) achieve productivity because they use symbol manipulation to process or create compositional (or combinatorial) structures.

Symbol manipulation depends on the ability to make copies of symbols and to transport them to other locations. As described by Newell [22, page 74]: "The symbol token is the device in the medium that determines where to go outside the local region to obtain more structure. The process has two phases: first, the opening of access to the distal structure that is needed; and second, the retrieval (transport) of that structure from its distal location to the local site, so it can actually affect the processing. (...) Thus, when processing "The cat is on the mat" (which is itself a physical structure of some sort) the local computation at some point encounters "cat"; it must go from "cat" to a body of (encoded) knowledge associated with "cat" and bring back something that represents that a cat is being referred to, that the word "cat" is a noun (and perhaps other possibilities), and so on."

Symbols can be used to access and retrieve information because they can be copied and transported. In the same way, symbols can be used to create combinatorial structures. In fact, making combinatorial structures with symbols is easy. This is why symbolic architectures excel in storing, processing and transporting huge amounts of information, ranging from tax returns to computer games. The capacity of symbolic architectures to store (represent) and process these forms of information far exceeds that of humans.

But interpreting information in a way that could produce meaningful answers or purposive actions is far more difficult with symbolic architectures. In part, this is due to the ungrounded nature of symbols. The ungrounded nature of symbols is a direct consequence of using symbols to access and retrieve information, as described by Newell. When a symbol token is copied and transported from one location to another, all its relations and associations at the first location are lost. For example, the perceptual information related to the concept *cat* is lost when the symbol token for *cat* is copied and transported to a new location outside the location where perceptual information is processed. At the new location, the perceptual information related to cats is not directly available. Indeed, as Newell noted, symbols are used to escape the limited information that can be stored at one site. So, when a symbol is used to transport information to other locations, at least some of the information at the original site is not transported.

The ungrounded nature of symbol tokens has consequences for processing. Because different kinds of information related to a concept are stored and processed at different locations, they can be related to each other only by an active decision to gain access to other locations, to retrieve the information needed. This raises the question of who (or what) in the architecture makes these decisions, and on the basis of what information. Furthermore, given that it takes time to search and retrieve information, there are limits on the amount of information that can be retrieved and the frequency with which information can be renewed.

So, when a symbol needs to be interpreted, not all of its semantic information is directly available, and the process to obtain that information is very time consuming. And this process needs to be initiated by some cognitive agent. Furthermore, implicit information related to concepts (e.g., patterns of motor behavior) cannot be transported to other sites in the architecture.

7.1. Grounded Representations. In contrast to symbolic representations, conceptual representations in human cognition are grounded in experiences (perception, action, emotion) and (conceptual) relations (e.g., [23, 24]). The forms of representation discussed in Section 6.2 are all grounded in this way. For example, grounding of visual representations begins with the retinotopic (topographic) representations in the early visual cortex. Likewise, motor representations are grounded because they are based on the frame of reference determined by the body parts that are involved in the execution of a movement. An arbitrary symbol is not grounded in this way.

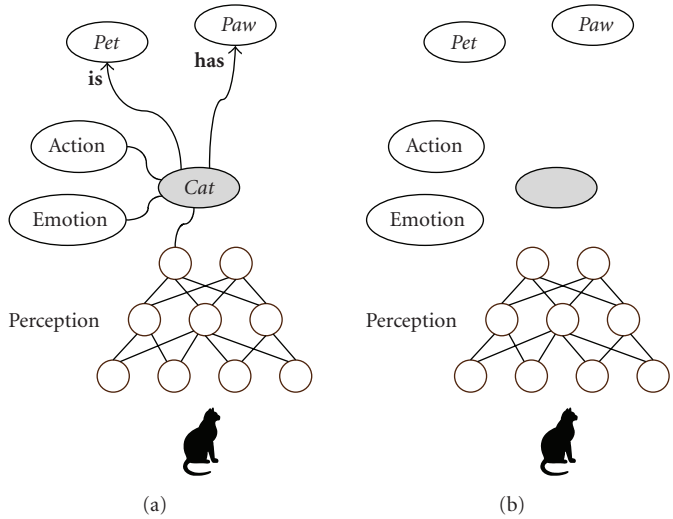

Figure 1: (a) illustration of the grounded structure of the concept *cat*. The circles and ovals represent populations of neurons. The central population labeled *cat* can be used to bind the grounded representation to combinatorial structures. (b) without the overall connection structure, the central population no longer forms a representation of the concept *cat*.

The consequence of grounding, however, is that representations cannot be copied and transported elsewhere. Instead, they consists of a network structure distributed over the cortex (and other brain areas). An illustration is given in Figure 1, which illustrates the grounded structure of the concept *cat*.

The grounded representation of *cat* interconnects all features related to cats. It interconnects all perceptual information about cats with action processes related to cats (e.g., the embodied experience of stroking a cat, or the ability to pronounce the word *cat*), and emotional content associated with cats. Other information associated or related to cats is also included in the grounded representation, such as the (negative) association between cats and dogs and the semantic information that a cat is a pet or has paws.

It is clear that a representation of this kind develops over time. It is in fact the grounded nature of the representation that allows this to happen. For example, the network labeled "perception" indicates that networks located in the visual cortex learn to identify cats or learn to categorize them as animals. In the process of learning to identify or categorize cats they will modify their connection structure, by growing new connections or synapses or by changing the synaptic efficacies. Other networks will be located in the auditory cortex, or in the motor cortex or in parts of the brain related to emotions. For these networks as well, learning about cats results in a modified network structure. Precisely because these networks remain located in their respective parts of the cortex, learning can be a gradual and continuous process. Moreover, even though these networks are located in different brain areas, connections can develop over time between them because their positions relative to each other remain stable as well.

The grounded network structure for *cat* illustrates why grounded concepts are different from symbols. There is no well designated neural structure like a symbol that can be copied or transported. When the conceptual representation of *cat* is embodied in a network structure as illustrated in Figure 1, it is difficult to see what should be copied to represent *cat* in sentences like these.

For example, the grey oval in Figure 1, labeled *cat*, plays an important role in the grounded representation of the concept *cat*. It represents a central neural population that interconnects the neural structures that represent and process information related to cats. However, it would be wrong to see this central neural population itself as a neural representation of *cat* that could be copied and transported like a symbol. As Figure 1 (b) illustrates, the representational value of the central neural population labeled *cat* derives entirely from the network structure of which it is a part. When the connections between this central neural population and the other networks and neural populations in the structure of *cat* are disrupted, the central neural population no longer constitutes a representation of the concept *cat*. For example, because it is no longer activated by the perceptual networks that identify cats. So, when the internal network structure of the central neural population (or its pattern of activation) is copied and transported, the copy of the central neural population is separated from the network structure that represents *cat*. In this way, it has lost its grounding in perception, emotion, action, associations and relations.

7.2. Grounded Representations and Productivity. Making combinatorial structures with symbols is easy. All that is required is to make copies of the symbols (e.g., words) needed and to paste them into the combinatorial structure as required. This, of course, is the way how computers operate and how they are very successful in storing and processing large amounts of data. But as noted above, semantic interpretation is much more difficult in this way, as is the binding with more implicit forms of information

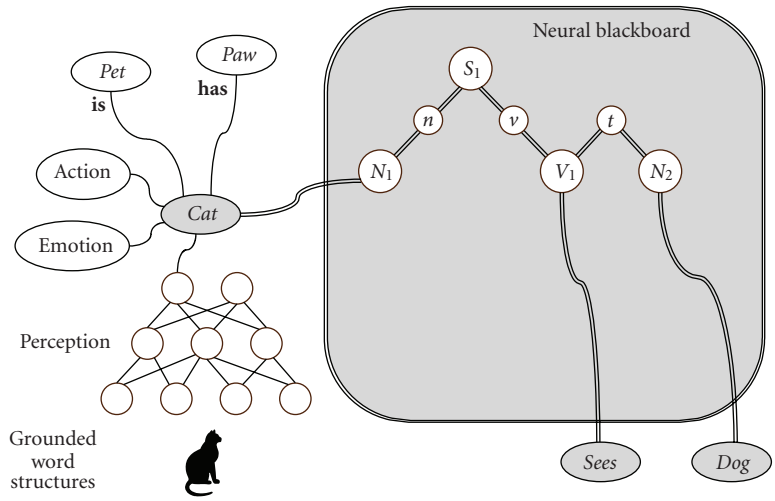

Figure 2: Illustration of the combinatorial structure *The cat sees the dog* (ignoring *the*), with grounded representations for the words. The circles in the neural blackboard represent populations and circuits of neurons. The double line connections represent conditional connections. (N, n = noun; S = sentence; t = theme; V, v = verb.)

storing found in embodied cognition. Yet, grounding representations and at the same time providing the ability to create novel combinatorial structures with these representations is a challenge, which the human brain seems to have solved.

At face value, there seems to be a tension between the grounded nature of human cognition and its productivity. The grounded nature of cognition depends on structures as illustrated in Figure 1. At a given moment, they consist of a fixed network structure distributed over one or more brain areas (depending on the nature of the concept). Over time, they can be modified by learning or development, but during any specific instance of information processing they remain stable and fixed.

But productivity requires that new combinatorial structures can be created and processed on the fly. For, as noted above, humans can understand and (potentially) produce in the order of 10^{20} (meaningful) sentences or more. Because this numbers exceeds the lifetime of the universe in seconds, it precludes that these sentences are somehow encoded in the brain by learning or genetic coding. Thus, most of the sentences humans can understand are novel combinatorial structures (based on familiar words), never heard or seen before. The ability to create or process these novel combinatorial structures was a main motivation for the claim that human cognition depends on symbolic architectures (e.g., [25]).

Figure 2 illustrates that grounded representations of the words *cat*, *sees* and *dog* can be used to create a combinatorial (compositional) structure of the sentence *The cat sees the dog* (ignoring *the*). The structure is created by forming temporal interconnections between the grounded representations of *cat*, *sees*, and *dog* in a "neural blackboard architecture" for sentence structure [13]. The neural blackboard consists of neural structures that represent syntactical type information (or "structure assemblies") such as structure assemblies for sentence (S_1), noun phrase (here, N_1 and N_2) and verb phrase (V_1). In the process of creating a sentence structure, the structure assemblies are temporarily connected (bound) to word structures of the same syntactical type. For example, *cat* and *dog* are bound to the noun phrase structure assemblies N_1 and N_2, respectively. In turn, the structure assemblies are temporarily bound to each other, in accordance with the sentence structure. So, *cat* is bound to N_1, which is bound to S_1 as the subject of the sentence, and *sees* is bound to V_1, which is bound to S_1 as the main verb of the sentence. Furthermore, *dog* is bound to N_2, which is bound to V_1 as its theme (object).

Figure 3 illustrates the neural structures involved in the representation of the sentence *cat sees dog* in more detail. To simplify matters, I have used the basic sentence structure in which the noun *cat* is connected directly to the verb *sees* as its agent. This kind of sentence structure is characteristic of a protolanguage [26], which later on develops into the more elaborate structure illustrated in Figure 2 (here, *cat* is the subject of the sentence, instead of just the agent of *sees*).

Figure 3(a) illustrates the structure of *cat sees dog*. The ovals are the grounded word structures, as in Figure 2. They are connected to their structure assemblies with memory circuits. The structure assemblies have an internal structure. For example, a noun phrase structure consists of a main part (e.g., N_1) and subparts, such as a part for agent (a) and one for theme (t). Subparts are connected to their main parts by gating circuits. In turn, similar subparts (or "subassemblies") of different structure assemblies are connected to each other by memory circuits. In this way, N_1 and V_1 are connected with their agent subassemblies and V_1 and N_2 are connected with their theme subassemblies. This represents that *cat* is the agent of *sees* and *dog* is its theme.

The structure assemblies (main parts and subparts alike) consists of pools or "populations" of neurons. So, each circle in Figure 3 represents a population. The neurons in a population are strongly interconnected, which entails that a population behaves as a unity, and its behavior can be modeled with population dynamics [13]. Furthermore,

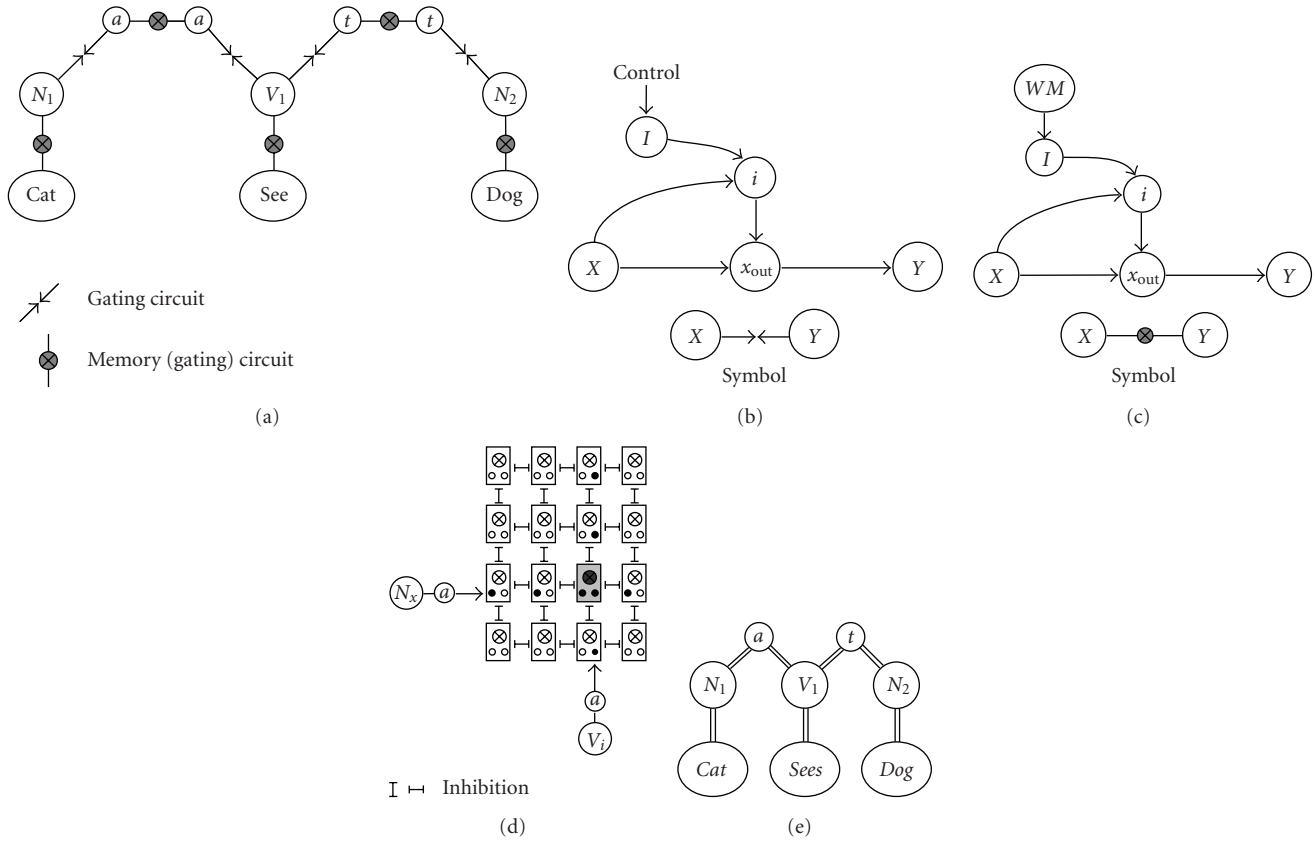

Figure 3: Illustration of the detailed neural structures involved in a sentence representation as illustrated in Figure 2. Ovals represent grounded word structures. The oval *WM* represents a working memory population, that remains active for a while after being activated. Circles represent populations of neurons. *I* and *i* are inhibitory neuron populations. The other ones are excitatory populations. (*a* = agent; *N* = noun; *t* = theme; *V* = verb.)

a population can retain activation for a while, due to the reverberation of activity within the population [27].

Figure 3(b) illustrates a gating circuit between two populations (*X* and *Y*). It consists of a disinhibition circuit. Activation can flow from *X* to *Y* when a control circuit activates population *I*, which in turn inhibits population *i*. The combination of gating circuits from *X* to *Y* and from *Y* to *X* is represented by the symbol illustrated in Figure 3(b). Gating circuits provide control of activation. They prevent that interconnected word structures form an associative structure, in which all word structures become automatically activated when one of them is active. Instead, activation from one word structure to another depends on specific control signals that activate specific gating circuits. In this way, information can be stored and retrieved in a precise manner. For example, the architecture can answer the question "What does the cat see?" or "Who sees the dog?" in this way [13].

Figure 3(c) illustrates a memory circuit between two populations (*X* and *Y*). It consists of a gating circuit that is activated by a working memory (*WM*) population. The *WM* population is activated when *X* and *Y* have been activated simultaneously (using another circuit not shown here [13]). So, the *WM* population stores the "memory" that *X* and *Y* have been activated simultaneously. Activation in the *WM* population consists of reverberating (or "delay") activity, which remains active for a while [27]. The combination of memory circuits from *X* to *Y* and from *Y* to *X* is represented by the symbol illustrated in Figure 3(c). When the *WM* population is active, activation can flow between *X* and *Y*. In this way, *X* and *Y* are "bound" into one population. Binding lasts as long as the *WM* population is active.

Bindings in the architecture are between subassemblies of the same kind (this is, in fact, also the case for the bindings between word assemblies and structures assemblies, although these subassemblies are ignored here). Figure 3(d) shows the connection matrix for binding between the agent subassemblies of noun phrase and verb phrase structure assemblies. All other subassembly bindings depend on a similar connection matrix. Arbitrary noun phrase and verb phrase structure assemblies can bind in this way. Binding occurs in a "neural column" that interconnects their respective subassemblies (agent subassemblies in this case). The neural column consists of the memory circuits needed for binding (and the circuit that activate the *WM* population). Neural columns for the same noun phrase or verb phrase structure assembly inhibit each other, which ensures that a noun phrase can bind to only one verb phrase structure assembly (and vice versa) with the same subassembly.

Figure 3(e) illustrates a "shorthand" representation of the entire connection structure of the sentence *cat sees dog* illustrated in Figure 3. When subassemblies are bound by memory circuits, they effectively merge into one population, so they are represented as one. The gating circuits, and the memory circuits between word and structure assemblies, are represented by double lines. The structure as represented in Figure 3(e) in fact consists of more than 100 populations, consisting of the populations that represent the structure assemblies and the populations found in the gating and memory circuits. To "see" these populations, one would have to "unwrap" the shorthand representation, inserting the connection matrices, gating and memory circuits and structure assemblies involved.

In the remainder of the paper, I will use the shorthand notion, as I have done in Figure 2. But the full structure is always implied, consisting of over 100 populations (substantially more for more complex sentences). So, for example, the circle labeled "*n*" in Figure 2 represents the "noun" subassemblies of the N_1 and S_1 structure assemblies, and the memory circuit that connects them. In this way, N_1 is bound to S_1 as its subject. Likewise, S_1 and V_1 are connected with their "verb" (*v*) subassemblies.

All bindings in this architecture are of a temporal nature. Binding is a dynamic process that activates specific connections in the architecture. The syntax populations (structure assemblies) play a crucial role in this process, because they allow these connections to be formed. For example, each word structure corresponding to a noun has connections to each noun phrase population in the architecture. However, as noted, these connections are not just associative connections, due to the neural (gating) circuits that control the flow of activation through the connection.

To make a connection active, its control circuit has to be activated. This is an essential feature of the architecture, because it provides control of activation, which is not possible in a purely associative connection structure. In this way, relations instead of just associations can be represented. Figure 1 also illustrates an example of relations. They consist of the conditional connections between the word structure of *cat* and the word structures of *pet* and *paw*. For example, the connection between *cat* and *pet* is conditional because it consists of a circuit that can be activated by a query of the form *cat is*. The **is** part of this query activates the circuit connection between *cat* and *pet*, so that *pet* is activated as the answer to the query. Thus, in conditional connections the control of activation can be controlled. For example, the **is** and **has** labels in Figure 1 indicate that information of the kind *cat is* or *cat has* controls the flow of activation between the word structures.

In Figures 2 and 3, the connections in the neural blackboard and between the word structures and the blackboard are also conditional connections, in which flow of activation and binding are controlled by circuits that parse the syntactic structure of the sentence. These circuits, for example, detect (simply stated) that *cat* is a noun and that it is the subject of the sentence *cat sees dog*. However, the specific details of the control and parsing processes that allow these temporal connections to be formed are not the main focus of this article. Details can be found in [9]. Here, I will focus on the general characteristics that are required by any architecture that combines grounded representations in a productive way. Understanding these general features is important for the interaction between AI and neuroscience.

7.3. Characteristics of Grounded Architectures. The first characteristic is the grounded nature of representations in combinatorial structures. In Figures 2 and 3, the representations of *cat*, *sees*, and *dog* remain grounded in the whole binding process. But the structure of the sentence is compositional. The syntax populations (structure assemblies) play a crucial role in this process, because they allow temporal connections to be formed between grounded word representations. For example, the productivity of language requires that we can form a relation between an arbitrary verb and an arbitrary noun as its subject. But we can hardly assume that all word structures for nouns are connected with all word structures for verbs, certainly not for noun verb combinations that are novel. Yet, we can assume that there are connections between words structures for nouns and a limited set of noun phrase populations, and that there are connections between words structures for verbs and a limited set of verb phrase populations. And we can assume that there are connections between noun phrase and verb phrase populations. So, using the indirect link provided by syntax populations we can create new (temporal) connections between arbitrary noun and verbs, and temporal connections between words of other syntactic types as well.

The second characteristic is the use of conditional and temporal connections in the architecture. Conditional connections provide a control of the flow of activation in connections. This control of activation is necessary to encode relational information. By controlling the flow of activation the architecture can answer specific queries such as *what does the cat see?* or *who sees the dog?*. Without such control of activation, only associations between word (concept) structures could be formed. But when connections are conditional and temporal (i.e., their activation is temporal), arbitrary and novel combinations can be formed in the same architecture (see [13]).

The third characteristic is the ability to create combinatorial structures in which the same grounded representation is used more than once. Because grounded representations cannot be copied, another solution is needed to solve this problem of multiple instantiations, that is, the "problem of two" [9]. Figure 4 illustrates this solution with the sentences *The cat sees the dog* and *The dog sees the cat* (ignoring *the*). The combinatorial structures of these two sentences can be stored simultaneously in the blackboard architecture, without making copies of the representations for *cat*, *sees* and *dog*. Furthermore, *cat* and *dog* have different syntactic roles in the two sentences.

Figure 4 illustrates that the syntax populations eliminate the need for copying representations to form sentences. Instead of making a copy, the grounded representation of *cat* is connected to N_1 in the sentence *cat sees dog* and to N_4 in the sentence *dog sees cat*. Because N_1 is connected to S_1,

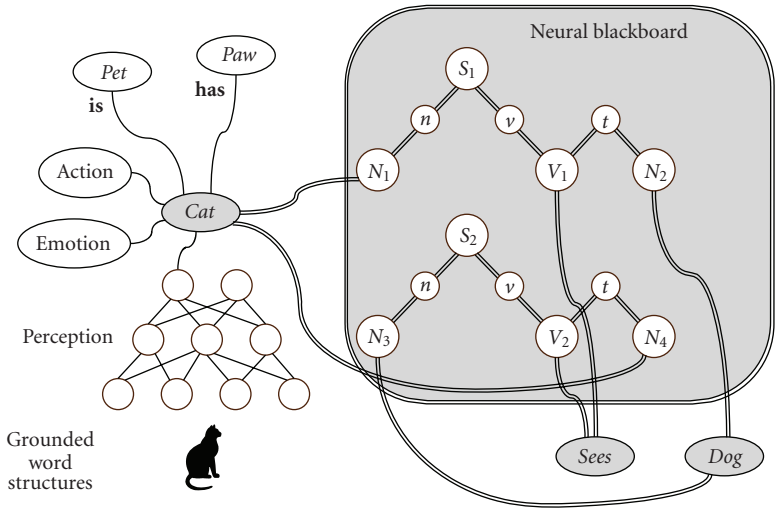

Figure 4: Illustration of the combinatorial structures of *The cat sees the dog* and *The dog sees the cat* (ignoring *the*), with grounded representations for the words. The circles in the neural blackboard represent populations and circuits of neurons. The double line connections represent conditional connections. (N, n = noun; S = sentence; t = theme; V, v = verb.)

cat is the subject in the sentence *cat sees dog*. It is the theme (object) in the sentence *dog sees cat*, because N_4 is connected to V_2 as its theme. The multiple binding of the grounded representations *dog and sees* proceeds in a similar way.

The fourth characteristic concerns the (often sequential) control of activation in the architecture. As I noted above, the conditional connections provide the ability to control the flow of activation within the architecture. Without this control, the architecture cannot represent and process combinatorial structures and relations. Control of activation results from neural circuits that interact with the combinatorial structures. Examples of control circuits can be found in [13, 28].

Figure 5 illustrates how these control circuits can affect and regulate the dynamics in the architecture, and with it the ability to process and produce information. With control of activation, the architecture can answer specific queries like *what does the cat see?* (or *cat sees?,* for short). The query *cat sees?* activates the grounded representations *cat* and *sees*. When the sentences *cat sees d*og and *dog sees cat* are stored in the blackboard, *cat* activates N_1 and N_4, because it is temporarily bound with these syntax populations. Likewise, *sees* activates V_1 and V_2.

But the query *cat sees?* also provides the information that *cat* is the subject of a verb. Using this information, control circuits can activate the conditional connections between subject syntax populations. In Figure 5 these are the connections between N_1 and S_1 and between N_3 and S_2. Because *cat* has activated N_1, but not N_3, N_1 activates S_1. Notice that the activation of N_4 by *cat* has no effect here, because N_4 is bound to V_2 as its theme (t), and these conditional connections are not activated by the query (yet). Because *cat* is the subject of a verb (*sees*), this information can be used to activate the conditional connections between the S_i and V_j populations in the architecture. Because S_1 is the only active S_i population, this results in the activation of V_1 by S_1.

At this point, a fifth characteristic of grounded cognition emerges: the importance of dynamics. Figure 5 shows why dynamics is important. Because *sees* is grounded, the query *cat sees?* has activated all V_j populations bound to *sees*, here V_1 and V_2. This would block the answer to the query, because that consists of activating the theme of V_1 but not the theme of V_2. However, due to the process described above, S_1 also activates V_1. Because populations of the same nature compete in the architecture (by inhibition), V_1 wins the competition with V_2.

When V_1 has won the competition with the other V_j populations, the query can be answered. The query *cat sees?* asks for the theme of the verb for which *cat* is the subject. That is, its asks for the theme of a syntax population bound to *sees*. After the competition, V_1 has emerged as the winning syntax population bound to that verb, so the query asks for the theme of V_1. It can do so by activating the conditional connections between V_1 and N_2 (see [9]). This will result in the activation of N_2 and with that of *dog* as the answer to the query.

The sequential nature of control illustrated in Figure 5 resembles that of control of movement. Executing a particular movement usually consists of sequential activation of a set of muscles. For example, when we swing an arm back and forth, its muscles have to be activated and deactivated in the correct sequence. More complex movement patterns like dancing or piano playing require elaborate sequential control of muscles being activated and deactivated. The motor programs for these movement patterns could in fact be a basis for the development of grounded representations. After all, muscles are "grounded" by nature. That is, we have just one set of muscles that we use to make specific movement sequence.

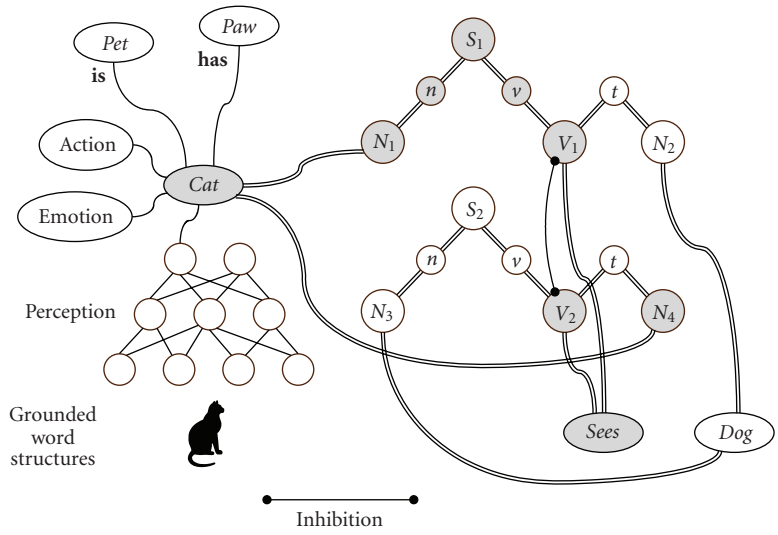

Figure 5: Illustration of the combinatorial structures of *The cat sees the dog* and *The dog sees the cat* (ignoring *the*), with grounded representations for the words. The circles in the neural blackboard represent populations and circuits of neurons. The grey nodes represent activate populations initiated by the query *cat sees?*. The double line connections represent conditional connections. (N, n = noun; S = sentence; t = theme; V, v = verb.)

7.4. Blackboard Architectures for Cognitive Processing. The combination of productivity and grounding requires certain architectures in which the grounded representations can be combined temporarily into combinatorial structures. The neural blackboard for sentence structure illustrated in Figures 2 and 3 is an illustration of such an architecture.

The neural blackboard illustrated in Figures 2, 3 and 4 provides the ability to form sentence structures. But words, for example, also have a phonological structure, and these structures are productive (combinatorial) as well. So, words would also be a part of a phonological neural blackboard. Words (concepts) can be used in reasoning processes based on sentence structures, which would require a specific blackboard architecture as well [13]. But words could also be a part of nonsentence like sequences, which could be used for other specific forms of reasoning [29]. Because the sentence blackboard is not suited for these sequences, a specific sequence blackboard is required as well.

Thus, grounded conceptual representations will be embedded in neural blackboards for sentence structure, phonological structure, sequences and reasoning processes, and potentially other blackboards as well. One might argue that this is overly complex. But complexity is needed to account for human cognition. Complexity is hidden in symbol manipulation as well. For example, when a specific symbol manipulation process is executed on a computer, a lot of its complexity is hidden in the underlying machinery provided by the computer. As a model of cognition, this machinery has to be assumed as a part of the model.

Furthermore, the embedding of representations in different blackboards is a direct consequence of the grounded nature of representations. Because these representations always remain "in situ", they have to be connected to architectures like blackboards to form combinatorial structures and to execute processes on the basis of these structures. In fact, the grounded representations form the link between the different blackboard architectures. When processes occur in one blackboard, the grounded representation can also induce processes in the other blackboards, which could in turn influence the process in the first blackboard. In this way, an interaction occurs between local information embodied in specific blackboards and global information embodied in grounded representations.

Viewed in this way, architectures of grounded cognition reverse the relation between control and representation as found in symbolic architectures of cognition. In the latter, resembling the digital computer, control is provided by a central "fixed" entity (e.g., the CPU) and representations move around in the architecture, when they are copied and transported. In grounded cognition, however, the representations are "fixed", whereas control moves around within and between blackboards.

8. Research Directions: Searching for Grounded Architectures of Cognition

The analysis given above suggests that cognition on the level of human cognition arises from the interaction between grounded representations and productive (blackboard) architectures. If so, these grounded architectures (for short) would have to be instantiated in the brain. This raises the question of how one could demonstrate that these architectures exist, and how their properties could be studied.

Empirical techniques such as electrodes, EEG (electroencephalogram) and fMRI (functional magnetic resonance imaging) are used to study "cognition in the brain". Each of these techniques provides valuable information about how the brain instantiates cognition. But each of them

is also limited. EEG provides information about groups of neurons (typically in the millions), for the most part located at the surface of the cortex. It's temporal resolution is very high, whereas its spatial resolution is relatively low. Functional MRI provides better spatial resolution (although not on the level of the neuronal circuits as found in cortical columns), but it's temporal resolution is too low to capture the dynamics of cognition.

Electrodes, inserted in the cortex, have the best spatial and temporal resolution. But the number of electrodes that can be inserted is limited relative to the number of neurons involved in a cognitive process. Moreover, it's use in humans is restricted to specific cases that arise when humans need brain surgery for medical reasons (e.g., [8]). A rigorous use as in animal experiments is excluded with humans for obvious ethical reasons. But the consequence of that is that detailed theories and models of human cognitions could never be tested empirically in detail.

It is important to emphasize this point, because it entails an additional difficulty that the study of human cognition faces. A scientific requirement of theories and models is that they can be tested empirically. Sometimes, theories and models cannot be tested (in full) because they are (partly) too vague or ambiguous. Such theories and models do not meet scientific standards in full. But in the case of human cognition, theories and models could be exact, detailed and unambiguous, but fail empirical testing due to ethical reasons. This is particularly true for the features of cognition that are specifically human. Detailed information is available for visual processing, for example, because we have animal models to test and investigate vision. But animal models are missing for language, planning, reasoning and other more exclusively human forms of cognition.

Perhaps the only way to test theories and models for these features of human cognition is large-scale brain modelling. Animal models could be of value because they provide the initial information and testing for simulating cortical columns, areas and pathways. Given the uniform nature of the cortex, between and within animals, these simulations could form the basis for cortical models of cognitive processes that are more specifically human. As suggested here, these models would consist of grounded architectures. These architectures require more than the simulation of cortical structures suited for animal cognition. For example, specific connection structures are needed to create blackboard architectures [9]. So, simulations need to investigate how these connection structures can be formed with cortical columns, or how other connection structures can be formed with cortical columns that have the same functional abilities.

In this process, AI would take a leading role, because it can develop detailed models of cognitive processes based on neural architectures. These models could then be used as a target for cortical simulations. That is, with cortical simulations it could be investigated whether and how the neural models developed by AI can be instantiated with the cortical building blocks found in the brain. In turn, these cortical simulations could be investigated by deriving virtual measurements from them, resembling electrode, EEG and even fMRI measurements. The latter could then be compared with measurements derived from actual brains.

The role of AI in this process is to analyze the mechanisms that can produce high-level processes of human cognition, and to develop neural instantiations for these mechanisms, such as the neural blackboard architectures discussed in the previous section. Neuroscience would provide the detailed information about the cortical building blocks, as discussed earlier. Large-scale simulations would integrate and further develop these two lines of investigation. So, AI has an important role to play in this research. But AI may also benefit from it, because this research could also solve important issues concerning the nature and mechanisms of intelligence and cognition. I will briefly discuss some of them in the final section.

9. Investigating Deep Problems

A number of issues in the study of (human) cognition can be characterized as "deep" problems. They concern the very nature of human-level cognition, and they have been the topic of speculation from the very beginning of thinking about cognition. But they largely remain as problems to be solved. The lack of progress with these problems also has a clear negative effect on the development of artificial forms of intelligence. The solution of these problems is most likely to be found in the unique way in which the human brain produces cognition, and thus in the unique computational and cognitive features of the neural architectures in the brain. Motivation for this assumption is found in the fact that the human brain is the only known example of a system that produces (human-level) cognition. Investigating the neural architectures of cognition thus provides the possibility to study at least some of these problems in a way that has not been available before. A few of these problems can be singled out.

9.1. Conceptual Structure (Meaning). Arbitrary symbols, gestures or sounds can be used to convey meaning, such as words and sentences in language. The question is how arbitrary symbols and sounds acquire meaning, what the nature (structure) of their meaning is, and how they succeed in conveying their meaning. An indication of the profound nature of these questions is the fact that meaning is one of the major problems in automatic language translation. Neuroimaging research has already demonstrated that there are relations between the neural representation of certain words and sensory-motor representations in the brain (e.g., action verbs activate parts of the motor cortex that are involved in the actions these verbs denote). Given these relations, it can be assumed that the nature and development of certain conceptual representations in the brain are related to the nature and development of sensory representations (e.g., sensory categorizations), motor representations, or transformations between representations. Thus, the study of sensory-motor representations and their transformations in neural architectures (as outlined above) could also be used to study the nature and development of those conceptual

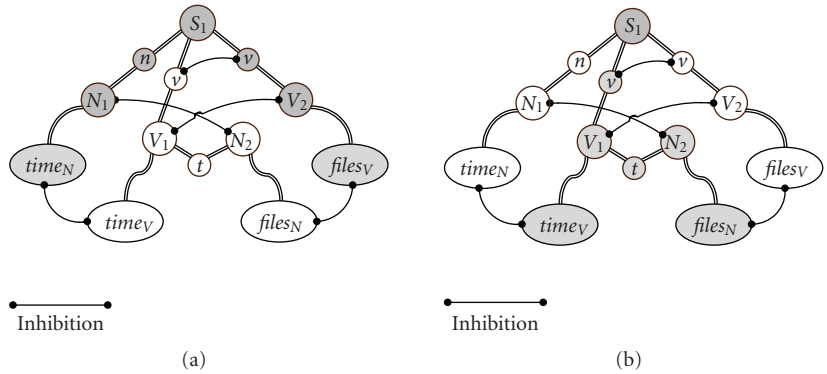

Figure 6: Competing neural blackboard structures for *time flies*. In (a), the competition results in $time_N flies_V$. In (b) the competition results in $time_V flies_N$. The ovals and circles represent populations as in Figures 2 and 3. Grey circles and ovals are active. (N, n = noun; S = sentence; t = theme; V, v = verb.)

structures that are related to sensory representations (e.g., nouns or adjectives), motor representations (e.g., verbs), or transformations (e.g., prepositions, [9]).

9.2. Selection of Information (Resolution of Ambiguity). Information is often ambiguous. A good illustration is given by language. Almost all words in language have multiple meanings. Consequently, sentences are very often ambiguous. For example, in the sentence *Time flies like an arrow*, the word *time* can be a noun, a verb, or even an adjective (i.e., *time flies* as in *fire flies*). Furthermore, the word *flies* can be a verb or a (plural) noun and the word *like* can be a verb or an adverb. Each of these choices provide a different interpretation for this sentence, for which at least five different interpretations are possible [19]. Artificial (computer) programs for sentence interpretation and translation have substantial difficulties in handling these forms of ambiguity.

Ambiguities are common in language and cognition in general, but humans often do not notice them [30, 31]. This is also the case for the sentence *Time flies like an arrow*. The usual interpretation of this sentence is in terms of a metaphor, that states that time changes very fast. Humans usually end up with this (one) interpretation, but a computer program of sentence analysis (based on symbol manipulation) gave all five interpretations [19]. The fact that humans can operate remarkably well with ambiguous sentences indicates that they have the ability to select the relevant or intended meaning from the ambiguous information they receive.

The difficulty of artificial intelligence systems to select relevant information has been another major problem in their development (sometimes referred to as the frame problem). Selecting relevant information is in particular a problem for generative (rule-based) processing. It is in fact the downside of the productivity of this form of processing. With generative processing, too many possibilities to be explored are often produced in a given situation. In contrast, associative structures such as neural assemblies are very suited for selecting relevant information. For example, when information in a neural assembly is partly activated, the assembly will reactivate all related information as well. The ability to select relevant information in human cognition could thus result from a combination of generative and associative processing. The development of grounded neural architectures of cognition, in which neural assemblies are combined with generative processing in neural blackboard architectures, as illustrated above, provides a way to investigate this possibility.

Figure 6 illustrates how ambiguity resolution could occur in a neural architecture of grounded cognition. In the architecture, dynamical interactions can occur between sentence structures [9]. Similar interactions can also influence the binding process, that is, the process by which a sentence structure is formed [19]. Figure 6 shows the competing sentence structures of *time flies*. The word *time* activates two grounded (word) structures, one for *time* as a noun ($time_N$) and one for *time* as a verb ($time_V$). In the same way, *flies* activates $flies_N$ and $flies_V$.

Initially each of the word structures binds to corresponding syntax populations, such as N_1 and V_1. These syntax populations then form competing sentence structures. One is the sentence structure for $time_N flies_V$ (the grey nodes in Figure 6(a)). Here, $time_N$ is the subject of the sentence and $flies_V$ is the main verb. The other is the sentence structure for $time_V flies_N$ (the grey nodes in Figure 6(b)). Here, $flies_N$ is the theme (t) of the verb $time_N$.

In the architecture, there is a dynamic competition between the sentence structures and between word structures. In particular, the word structures for $time_N$ and for $time_V$, and those for $flies_N$ and $flies_V$ inhibit each other. This competition implements the constraint that a word can have only one interpretation at the same time in a sentence structure. Between the sentence structures there is a competition (inhibition) between the circuits that activate conditional connections of the same kind (in Figure 6 those for the verb connections), and inhibition between similar syntax populations (e.g., between the noun phrases N_1 and N_2 and between the verb phrases V_1 and V_2).

The outcome of the competition is either the structure illustrated with the grey nodes in Figure 6(a), or the structure with the grey nodes in Figure 6(b). The competition is

resolved when there is a clear advantage for one of the competing structures [13, 28]. In Figure 6, an advantage for one of the sentence structures can arise from the fact that the interpretation of *time* as a noun is more frequent than the interpretation of *time* as a verb. In that case, the activation of $time_N$ will be stronger than that of $time_V$, so that the first inhibits the second. Then, $flies_V$ inhibits $flies_N$, because $flies_V$ is activated by $time_N$ through the sentence structure, whereas N_2 is inhibited by N_1 (this inhibition becomes stronger with increasing activation of $time_N$). In this way, the grey structure in Figure 6(a) remains as the active structure to which the rest of the sentence, *like an arrow*, binds.

The competition in Figure 6 illustrates why grounded representations are important, and why they have to remain grounded in combinatorial structures. The competition that solves the ambiguity of *time flies*, for example, results from the interaction between the structures of $time_N$ and $time_V$. The assumption is that $time_N$ wins this competition because it is used more frequently than $time_V$ in natural language. Due to the grounded nature of representations, the more frequent use of $time_N$ will affect the grounded representation of $time_N$ directly, because this representation is always used to represent $time_N$. Furthermore, the difference between $time_N$ and $time_V$ is found only in sentence contexts, thus in combinatorial structures. So, when grounded representations remain grounded in combinatorial structures, the more frequently used type of combinatorial structure can influence the grounded structures involved directly.

9.3. Learning and Development. This is a topic of extensive research, which is very important for understanding cognition. One problem concerning learning and development perhaps stands out. It concerns the difference between associative versus generative (rule-based) processing, which in turn relates to the age-old debate between nature and nurture.

Associative processing plays an important role in human cognition. Examples are the neural assemblies proposed by Hebb [11]. Furthermore, the learning mechanisms discovered in the brain (e.g., long-term potentiation) concern the forming of new associations. Thus associative processing gives an account of the development of cognition (nurture). Examples are the associations that can develop within and between grounded conceptual representations. However, generative processing is needed for the productivity of cognition. But the development (learning) of generative processing is difficult to account for, which has led to the assumption that the basic principles of generative processing are innate (nature). Yet, these innate abilities develop only with proper stimulation (experience).

The problem thus concerns the question of what features of generative processing are innate, and how this innate ability develops on the basis of experience. If neural architectures of generative processing are adequately captured in a model, this problem could for the first time be addressed in a more experimental way, by using a backtracking procedure (reverse engineering). With this procedure one can simplify the known (fully developed) neural architecture and then investigate how the fully developed architecture can evolve from the more simplified version of it. This approach could be repeated in several steps, leading to a more and more elementary architecture as the basis of the fully developed architecture.

10. Conclusion

For the first time in history, it is possible to investigate the neural mechanisms that produce human cognition. It can be done because the experimental methods and techniques are now available to investigate the structure of the brain, because the theoretical knowledge is available that provides the possibility of a theoretical analysis of neural mechanisms of cognition, and because the computer power is now available that provides the possibility of large-scale simulations and numerical analyses of these mechanisms.

However, the complexity of the brain, and the cognitive processes it produces, entails that integrated multidisciplinary expertise is needed to combine these lines of research. The computational perspective on neurocognition, aimed at understanding how the neural dynamics and neural mechanisms of the brain produce cognition, can play a fundamental role in this respect, because it focuses on the ultimate aim of neurocognition [1]. So, AI has an important role to play in this process.

But AI can also benefit from it, because a detailed analysis of how the brain produces cognition could provide important information about the nature of cognition itself. Here, I have argued that understanding the neural basis of cognition could reveal important characteristics of its grounded nature. For example, combinatorial structures can be created with grounded representations, but not all structures are equally feasible [13]. And, as illustrated in Figure 6, the combinatorial structures formed are influenced by dynamics, which provides additional constraints on the ability to create combinatorial structures. The example given in Figure 6 show that these constraints prevent the excessive production of sentence interpretations, as found in systems with unlimited productivity based on symbol manipulation. But, on occasion, it can also result in misrepresentations, which is indeed found in human cognition as well.

The combination of grounding and productivity could solve a problem about cognition addressed by Fodor. Although he supported the computational view of cognition from its beginning, more recently Fodor has argued that a computational (symbol manipulation) account of cognition is incomplete [32]. In particular, because the computational processes provided by symbol manipulation are always local (as illustrated in the quote from Newell [22]). Local processing, in the view of Fodor, does not capture the global flexibility of cognition, which may be the most important feature of human cognition [32].

Grounded cognition, as presented here, is both local and global, and it is productive. Processes that occur within specific blackboards are local, but the grounded representations involved are global. The interaction between blackboards and grounded representations thus provides a basis for the productivity and global flexibility of cognition.

References

[1] M. S. Gazzaniga, "Preface," in *The Cognitive Neurosciences*, M. S. Gazzaniga, Ed., MIT Press, Cambridge, Mass, USA, 1995.

[2] V. Braitenberg, "Two views of the cerebral cortex," in *Brain Theory*, G. Palm and A. Aertsen, Eds., Springer, Berlin, Germany, 1986.

[3] V. Braitenberg and A. Schüz, *Anatomy of the Cortex: Statistics and Geometry*, Springer, Berlin, Germany, 1991.

[4] W. H. Calvin, "Cortical columns, modules, and Hebbian cell assemblies," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, G. Adelman, and P. H. Arbib, Eds., pp. 269–272, MIT Press, Cambridge, Mass, USA, 1995.

[5] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[6] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, 2007.

[7] K. Grill-Spector and R. Malach, "The human visual cortex," *Annual Review of Neuroscience*, vol. 27, pp. 649–677, 2004.

[8] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, "Invariant visual representation by single neurons in the human brain," *Nature*, vol. 435, no. 7045, pp. 1102–1107, 2005.

[9] R. Jackendoff, *Foundations of Language*, Oxford University Press, Oxford, UK, 2002.

[10] M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex*, Cambridge University Press, New York, NY, USA, 1991.

[11] D. O. Hebb, *The Organization of Behavior*, John Wiley & Sons, New York, NY, USA, 1949.

[12] E. Bienenstock, "Composition," in *Brain Theory: Biological Basis and Computational Theory of Vision*, A. Aertsen and V. Braitenberg, Eds., pp. 269–300, Elsevier, New York, NY, USA, 1996.

[13] F. van der Velde and M. de Kamps, "Neural blackboard architectures of combinatorial structures in cognition," *Behavioral and Brain Sciences*, vol. 29, no. 1, pp. 37–70, 2006.

[14] J. Frye, R. Ananthanarayanan, and D. S. Modha, "Towards real-time, mouse-scale cortical simulations," IBM Research Report RJ10404, 2007.

[15] H. Markram, "The blue brain project," *Nature Reviews Neuroscience*, vol. 7, no. 2, pp. 153–160, 2006.

[16] G. M. Shepherd, *Neurobiology*, Oxford University Press, Oxford, UK, 1983.

[17] R. J. Douglas and K. A. C. Martin, "Neocortex," in *The Synaptic Organization of the Brain*, G. M. Shepherd, Ed., pp. 389–438, Oxford University Press, Oxford, UK, 3rd edition, 1990.

[18] F. van der Velde and M. de Kamps, "From knowing what to knowing where: modeling object-based attention with feedback disinhibition of activation," *Journal of Cognitive Neuroscience*, vol. 13, no. 4, pp. 479–491, 2001.

[19] S. Pinker, *The Language Instinct*, Penguin, London, UK, 1994.

[20] G. A. Miller, *The Psychology of Communication*, Penguin, London, UK, 1967.

[21] J. R. Anderson, *The Architecture of Cognition*, Harvard University Press, Cambridge, Mass, USA, 1983.

[22] A. Newell, *Unified Theories of Cognition*, Harvard University Press, Cambridge, Mass, USA, 1990.

[23] S. Harnad, "The symbol grounding problem," in *Emergent Computation: Self-Organizing, Collective, and Cooperative Phenomena in Natural and Artificial Computing Networks*, S. Forrest, Ed., MIT Press, Cambridge, Mass, USA, 1991.

[24] L. W. Barsalou, "Perceptual symbol systems," *Behavioral and Brain Sciences*, vol. 22, no. 4, pp. 577–660, 1999.

[25] J. A. Fodor and Z. W. Pylyshyn, "Connectionism and cognitive architecture: a critical analysis," in *Connections and Symbols*, S. Pinker and J. Mehler, Eds., pp. 3–71, MIT Press, Cambridge, Mass, USA, 1988.

[26] W. H. Calvin and D. Bickerton, *Lingua ex Machina: Reconciling Darwin and Chomsky with the Human Brain*, MIT Press, Cambridge, Mass, USA, 2000.

[27] D. J. Amit, "The Hebbian paradigm reintegrated: local reverberations as internal representations," *Behavioral and Brain Sciences*, vol. 18, no. 4, pp. 617–657, 1995.

[28] F. van der Velde and M. de Kamps, "Learning of control in a neural architecture of grounded language processing," *Cognitive Systems Research*, vol. 11, no. 1, pp. 93–107, 2010.

[29] R. F. Hadley, "The problem of rapid variable creation," *Neural Computation*, vol. 21, no. 2, pp. 510–532, 2009.

[30] B. J. Baars and S. Franklin, "How conscious experience and working memory interact," *Trends in Cognitive Sciences*, vol. 7, no. 4, pp. 166–172, 2003.

[31] B. J. Baars, "Conscious cognition and blackboard architectures," *Behavioral and Brain Sciences*, vol. 29, no. 1, pp. 70–71, 2006.

[32] J. A. Fodor, *The Mind Doesn't Work That Way*, MIT Press, Cambridge, Mass, USA, 2000.