

THE TRANSITION FROM BIOLOGICAL TO CULTURAL EVOLUTION

WOLF SINGER

Introduction

The emergence of culture is with all likelihood the result of synergistic interactions among several evolutionary events, the most important consisting of an evolutionary refinement of brain functions. The use of tools, the formation of labour-sharing societies, the management of agriculture, and the development of language are above all other factors consequences of the increasing sophistication of the cognitive functions and motor skills provided by the brain of *homo sapiens sapiens*. Because our direct ancestors are all extinct it is extremely difficult to infer which aspects of brain development were actually decisive for the transition from apes to early hominids and finally culture-competent *sapiens-sapiens*. The only truly longitudinal data on the evolution of the human brain come from studies of fossil skulls. These analyses reveal a gradual increase in brain volume – but this notion is not particularly helpful because brain size alone, even if considered in relation to body weight, is only a poor correlate of functional sophistication. Thus, inferences on evolutionary changes in brain organisation have to rely on comparison of species that escaped extinction. However, and this is both interesting and unfortunate for studies of evolution, the surviving species all bifurcated from the line of our ancestors long before the gap that separates us from our nearest relatives, the apes. Therefore, only rather indirect inferences are possible.

What Makes the Difference?

The first question which arises is whether our brains differ from those of our ancestors who initiated cultural evolution, painted the walls of their

caves, and invented tools. The answer is yes and no. As far as genetically determined features are concerned, i.e. the molecular composition, the anatomical structures and the basic connectivity patterns, there cannot be any major differences because evolution is slow. This implies that our cave-dwelling ancestors were born with brains that must have had roughly the same inborn abilities as those of our babies. Hence, competences exceeding those of our ancestors must be attributed to the action of epigenetic factors, i.e. to experience-dependent modifications of postnatal brain development and to learning.

The two processes of epigenetic knowledge acquisition, experience dependent shaping of neuronal architectures during early development on the one hand, and learning on the other, differ at the neuronal level mainly with respect to the reversibility and the amplitude of the changes. Both are associated with lasting modifications of the interactions among neurons. During early development, experience can modify the architecture of neuronal connectivity by influencing the consolidation and disruption of newly formed pathways. Once these developmental processes decline, which is thought to occur around puberty, further modifications of functional architectures appear to be restricted to changes in the efficacy of the now consolidated repertoire of connections (see on this point Singer 1990, 1995). Thus, although the genetically determined blueprint of our brains is probably the same as that of our ancestors, our brains are likely to differ because of differences in the epigenetic shaping of fine structure.

This susceptibility of brain functions to undergoing epigenetic modifications is certainly a key factor in cultural evolution because it permits a highly efficient transmission of acquired abilities and knowledge from one generation to the next. It permits closure of the re-entry loop that couples collective achievements accomplished by interacting brains to the epigenetic shaping of ever more sophisticated functions of individual brains. Many of our cognitive abilities owe their differentiation to this collective learning process, and the reciprocity of the mechanism makes it difficult to determine which of our cognitive abilities have been the cause and which the result of cultural evolution. Despite this caveat our knowledge about the evolution of brains permits some educated guesses about which of the newly acquired neuronal functions might have actually triggered the onset of cultural evolution.

Concerning specific sensory or motor skills *homo sapiens sapiens* is a generalist. We perform reasonably well in many domains but for most of

them one can identify animals that outperform us. Still, if one distributed points for performance in the various sensory modalities and for basic motor skills we would with all likelihood come out as winners. It is unlikely, however, that this superiority in average performance is alone sufficient to account for the emergence of culture. Rather, our culture competence seems to result from the evolutionary development of certain cognitive functions that are unique to humans. One of these is probably our ability to generate abstract, symbolic meta-representations of cognitive contents by subjecting the results of first order cognitive operations iteratively to further cognitive processing of a higher order. This competence requires the ability to identify relations among the numerous distributed cognitive processes and to represent these relations. The results of such an iteration of cognitive operations are modality invariant and hence abstract descriptions of the outcome of first order cognitive processes. As these higher order descriptions are equivalent with an internal protocol that keeps track of the brain's own cognitive operations, they can be considered as the substrate of our ability to be aware of our own sensations and intentions as well as those of others. This awareness, in turn, is probably at the origin of our unique ability to use a symbolic communication system and to generate a theory of mind. We seem to be the only species that is capable of imagining the thoughts and emotions of the respective other when she or he is in a particular situation. We are the only species capable of entering into dialogues of the format "I know that you know that I know", or "I know that you know how I feel". Such dialogues permit not only a deeper understanding of the respective other but they also allow one to experience one's own cognitive functions in the reflection of the perceptions of the other. Thus, the ability to generate a theory of mind has probably been instrumental in the development of social interactions that shape our self-concepts and provide the basis for the experience that we are autonomous agents endowed with intentionality and free will. We experience these cultural constructs to be as real as pre-cultural realities and hence these social realities are likely to have as important a role in the epigenetic shaping of brain functions as the other environmental factors. This has deep consequences as it implies that cultural embedding influences the fine grained architecture of brains and hence part of the phenotype of the organism. Thus, cultural evolution is no longer constrained, as is biological evolution, by the inability to translate experience gathered by preceding generations into modifications of the phenotype of the offspring.

Neuronal Prerequisites for the Emergence of Culture

What do we know about the neuronal substrate that enables human brains to run protocols and generate meta-representations of their own performance, to realise what one might call an 'inner eye' function and to develop a theory of mind? What permits them to evaluate and to represent relationships between the distributed results of the various basic cognitive operations that occur in parallel and in relative isolation at lower levels of the brain? What could be the structure of the abstract meta-representations that integrate the modality-specific results provided by the various sensory systems and permit symbolic encoding of both external events and internal states?

One prerequisite for the generation of such higher order descriptions is a mechanism that allows for the binding of distributed first order processes. At the neuronal level the following requirements have to be fulfilled: 1) all computational results, both those of first and higher order processes, must be expressed in a common format to permit flexible recombination. 2) A versatile binding mechanism must be implemented that permits evaluation and representation of the relationships between the results of distributed computations. 3) Expanded storage capacities must be provided in order to maintain temporally dispersed contents in short-term buffers so that they are simultaneously available for the evaluation of relations and for binding. 4) Additional neuronal substrate needs to be provided for the generation of descriptions of higher order. 5) Effector systems are required that are sufficiently differentiated to permit the translation of the results of higher order computations into actions and the communication of the symbolic descriptions to other individuals.

When comparing our brains to those of non-human primates one is struck by their similarity and searches in vain for entirely novel structures that could account for the new, qualitatively different functions. At the macroscopic level, the only noticeable difference between our brains and those of non-human primates is an increase in the surface of the neocortex, and differences vanish nearly completely if one analyses the brains at the microscopic or molecular level. The internal organisation of the various brain structures, including the neocortex, is nearly identical, and the vast majority of the molecules expressed are the same. This leaves one with the conclusion that the new functions that distinguish *homo sapiens sapiens* from its nearest neighbour species must have been realised simply by the addition of further areas of the neocortex and/or by the rearrangement of connections among neocortical areas.

Comparative anatomy suggests that these additional areas differ from the more ancient areas in the way in which they are connected to sensory systems and effector organs. The new areas in the occipital, parietal and temporal lobes appear to be involved primarily in the refinement of sensory functions while the new areas in the frontal lobes subserve more executive functions such as action planning, short term storage and management of attention. The more recent sensory areas tend to receive their input not directly from the sensory periphery, as is the case for the more ancient sensory areas, but more indirectly via the latter. Moreover, the new areas tend to collect their input not from a single modality as the ancient sensory areas but from different modalities (Krubitzer, 1995, 1998). It is because of this peculiarity that the phylogenetically more recent areas which are topographically intercalated between the monomodal sensory areas have been addressed as association areas. The new areas in the frontal lobes are also more remote from the periphery than the more ancient motor centres. They tend to be connected to effector organs only indirectly via the ancient motor areas and receive most of their cortical input not from the primary sensory areas but from the more recent association areas. Thus, the connectivity of these phylogenetically recent cortical areas is compatible with the view that they re-evaluate and bind the distributed results of primary cognitive operations, and thereby provide the substrate for the generation of higher-order representations.

What remains puzzling, however, is the fact that all these different functions appear to rely always on the same computational algorithm. The intrinsic organisation of the neocortex is extremely complex but surprisingly stereotyped and monotonous. The laminar organisation, the various cell types, and the intrinsic connectivity differ only little between phylogenetically old and more recent cortical areas or between areas devoted to sensory and executive functions. Because the programme for the computational operations performed by neuronal networks is fully and exclusively determined by the architecture and coupling strength of connections, the structural homogeneity of the neocortex implies that the various regions perform more or less the same computations. This fascinating conclusion has recently received strong support from developmental studies in which inputs from the eye have been re-routed by surgical intervention to the auditory cortex, whereupon this piece of cortex developed exactly the same functional features as are normally characteristic for the visual cortex (Sharma *et al.*, 2000).

It appears, then, as if our brains owed their unique cognitive abilities simply to the iteration of processes realised by cortical architectures. All that seemed necessary for the development of new functions was apparently the addition of cortical areas which treat the output of the already existing areas in exactly the same way as these treat their input, which in lower animals comes mainly from the sensory periphery. In conclusion, the new cognitive abilities that distinguish humans from non-human primates seem to have emerged because evolution provided additional cortical areas which permitted reprocessing and binding of the results of first order processes and the generation of higher order, transmodal representations. Interestingly, the evolution of new cortical areas may not have required major changes in the genome as adding one more step of cell division to the division cycles of precursor cells of neocortical neurons can have dramatic effects on cortical cell numbers and hence cortical volume (Rakić, 1998).

The notion that neocortical modules process signals according to similar algorithms has the additional attractive implication that the results of their computations are likely to be encoded in the same format. Hence, they can be re-subjected in ever changing constellations to iterative processes of the same kind, thus generating representations of an increasingly higher order. Although this view is far from providing a mechanistic explanation for the emergence of phenomena such as phenomenal awareness, i.e. the ability to be aware of one's own sensations and actions, it provides at least an intuition of how brains can apply their cognitive abilities to some of their own processes, thereby creating descriptions of themselves and hypotheses about others.

In conclusion, it appears that any attempt to account for the emergence of those cognitive abilities that we consider instrumental for the evolution of culture needs to be based on an understanding of neocortical functions, and in particular of those that permit the binding of the results of distributed, primary cognitive operations into coherent, modality-independent and symbolic descriptions. Of primordial interest is, therefore, how contents are represented in cortical networks and how dynamic binding of these contents into meta-representations can be achieved.

The Structure of Representations

The hypothesis proposed here is that evolved brains use two complementary strategies in order to represent contents (see also Singer, 1995, 1999). The first strategy relies on individual neurons that are tuned to

respond selectively to particular constellations of input activity thereby establishing explicit representations of particular constellations of features. It is commonly held that the specificity of these neurons is brought about by selective convergence of input connections in hierarchically structured feed-forward architectures. This representational strategy allows for rapid processing and is ideally suited for the representation of frequently occurring stereotyped combinations of features. However, this strategy has several limitations. It is expensive in terms of the number of required neurons because it demands at least one neuron per object. Thus, it is not well suited to cope with the virtually infinite diversity of possible feature constellations encountered in real world objects. Moreover, this representational mode lacks systematicity which makes it difficult to encode relations between parts of the same object or semantic relations between different perceptual objects. A detailed discussion of the advantages and disadvantages of representing contents by individual smart neurons is to be found in Singer (1999), von der Malsburg (1999) and Gray (1999). The second strategy, according to this idea, consists of the temporary association of neurons into a functionally coherent assembly which as a whole represents a particular content whereby each of the participating neurons is tuned to one of the elementary features of the respective perceptual object. This representational strategy is more economical with respect to neuron numbers because a particular neuron can, at different times, participate in different assemblies just as a particular feature can be shared by many different perceptual objects. Moreover, this representational strategy allows for the rapid *de novo* representation of feature constellations that have never been experienced before. There are virtually no limits to the dynamic association of neurons in ever changing constellations, provided that the participating neurons are directly or indirectly connected. Thus, for the representation of highly complex and permanently changing contents this second strategy appears to be better suited than the first.

The meta-representations that result from iteration of cognitive operations are necessarily much richer in combinatorial complexity than the contents of first order processes. In addition, they must be highly dynamic because they need to be re-configured at the same pace as the contents of phenomenal awareness change. It appears then as if the second representational strategy that is based on the dynamic binding of neurons into functionally coherent assemblies would be more suitable for the implementation of higher order representations than the first strategy which relies on individual smart neurons. While the latter can readily be implemented in simple feed-forward networks and hence can be found also in the brains of

invertebrates, assembly coding requires neuronal architectures that permit in addition dynamic grouping of distributed responses through re-entry and self-organisation. This necessitates co-operative interactions among neurons and hence a complex network of reciprocal connections. It appears as if such architectures existed only in cortical structures, which may be one reason for the evolutionary success of the cerebral cortex.

Culture as a Prerequisite for the Emergence of Consciousness

The term 'consciousness' has a number of different connotations ranging from awareness of one's perceptions and sensations to self-awareness, the perception of oneself as an agent endowed with intentionality and free will. The first connotation of consciousness, phenomenal awareness, should in principle be tractable within neurobiological description systems because the problem can probably be reduced to the question of how neuronal representations are organised.

Brains that have the ability to bind the results of their distributed computational operations and to thereby generate metarepresentations of their own internal states can realise what one might call, an 'inner eye' function. They can become 'aware' of their own performance. The emergence of phenomenal awareness, which we readily grant also to higher mammals and primates, can thus be seen as a direct consequence of an evolutionary process; of a process that led to brain architectures which support the generation of metarepresentations by the iteration of elementary cognitive operations. The adaptive value of acquiring phenomenal awareness is obvious. It permits integration of the results of polymodal sensory processes with information stored in the memory and with the status of a value assigning system. Decisions for future acts can thus be based on a rich set of variables, they can be made dependent on what appears to the observer as 'reasoning' or 'internal deliberations'. Thus, brains endowed with such an 'inner eye' function can respond with more flexibility to changing conditions than brains that lack phenomenal awareness and are confined to reacting to stimuli without the option of further reflection and internal deliberation. This advantage may have been one reason for the evolution of brains capable of being aware of their own performance.

However, the other aspect of consciousness, the perception of one's self as an autonomous mental agent endowed with intentionality and free will seems to require explanations which transcend purely neurobiological reductionism. These aspects appear to have another ontological status as

the qualia of phenomenal awareness. We perceive ourselves as agents that are endowed with the freedom to decide, implying that the self is actually capable of controlling, by will, processes in the brain. We experience free will as an immaterial mental force that is capable of influencing the neuronal processes required for the execution of actions and, therefore, we perceive it as not derivable from the material processes in the brain. Hence, realities that we experience from our subjective first person perspective seem to differ ontologically from realities that we observe from the third person perspective when we analyse the functions of the brain. This coexistence of different description systems gives rise to several explanatory gaps. If one takes a dualistic stance one has to explain how an immaterial mental entity can interact with neuronal processes and induce action? If one adopts a monistic position, one has to explain how the nervous system, which is a deterministic material system that results from a continuous evolutionary and an equally continuous ontogenetic process can give rise to realities that are experienced as non-deterministic and immaterial. I propose that these explanatory gaps can be closed if one adopts a monistic position but includes as variables not only biological evolution and the physiology of individual brains but also cultural evolution.

I suggest that the experience that makes us believe that we are free is the result of cultural evolution, i.e. of interactions among brains that are sufficiently differentiated to be able to generate a theory of mind. Brains which possess phenomenal awareness and in addition have the ability to signal to one another and to comprehend that they are endowed with this capacity can mirror one another and attribute to themselves what they observe in the respective other. Many of the variables which determine the decisions and actions of the brain are not observable because they are not accessible to phenomenal awareness, and because they are not consciously perceived the most parsimonious interpretation of the driving force behind the unexplained causes for the actions of the respective other is the assignment of intentionality. Since we are ourselves also unaware of many of the variables which determine our own actions and since others attribute to us intentionality as well, and since this reflexive assignment accompanies us from early childhood throughout life, it appears quite plausible that we adopt this construct as a constitutive element of our self. The experience of an autonomous agent endowed with intentionality would thus be the result of reciprocal cognitive interactions among human beings. The experience of being free to act and as a consequence the conviction that one is responsible would then have to be considered as a product of social interactions

and the contents of this experience would have the ontological status of social realities, of cultural constructs.

The mechanisms that enable us to experience ourselves as endowed with mental capacities do, of course, reside in individual brains, but the contents of this experience are derived from social interactions. But why then should the experience of the intentional self be so obviously different from other experiences with social realities such as value systems, that we also derive from social interactions? One explanation could be that the interpersonal dialogue that nourishes the experience of the intentional self is already initiated during a very early developmental stage; a stage that precedes the maturation of episodic memory. If so, there would be no conscious record, no contextual embedding of the processes that led to the experience of the intentional self. Because of this amnesia these early experiences would lack causation, they would appear as timeless and detached from any real world context. In consequence, the contents of the experience of one's self, although acquired by learning, would be perceived as having transcendental qualities which resist reductionistic explanations.

LITERATURE

- Gray, C.M. 1999: 'The Temporal Correlation Hypothesis of Visual Feature Integration: Still Alive and Well', *Neuron* 24: 31-47.
- Krubitzer, L. 1995: 'The Organisation of Neocortex in Mammals: Are Species Differences Really so Different?', *Trends in Neurosciences* 18: 408-417.
- Krubitzer, L. 1998: 'Constructing the Neocortex: Influences on the Pattern of Organization in Mammals', in M.S. Gazzaniga and J.S. Altman (eds.), *Brain and Mind: Evolutionary Perspectives* (Strasbourg: HFSP Workshop V.), pp. 18-34.
- Rakić, P. 1998: 'Cortical Development and Evolution', in M.S. Gazzaniga and J.S. Altman (eds.), *Brain and Mind: Evolutionary Perspectives* (Strasbourg: HFSP Workshop V.), pp. 34-42.
- Sharma, J., Angelucci, A. and Sur, M. 2000: 'Induction of Visual Orientation Modules in Auditory Cortex', *Nature* 404: 841-847.
- Singer, W. 1990: 'The Formation of Cooperative Cell Assemblies in the Visual Cortex', *Journal Experimental Biology* 155: 177-197.
- Singer, W. 1995: 'Development and Plasticity of Cortical Processing Architectures', *Science* 270: 758-764.
- Singer, W. 1999: 'Neuronal Synchrony: A Versatile Code for the Definition of Relations?', *Neuron* 24: 49-65.
- Von der Malsburg, C. 1999: 'The What and Why of Binding: the Modeler's Perspective', *Neuron* 24: 95-104.