Contents

Special Issue

Unconscious Cognition and Evolution

Guest Editors: David L. Smith/Robert E. Haskell

David L. Smith/Robert E. Haskell

Preface

THIS SPECIAL ISSUE CONTAINS SELECTED PAPERS FROM the *1st International Conference of the New England Institute of Cognitive Science and Evolutionary Psychology* (NEI, http://www.une.edu/nei) on Unconscious Cognition and Evolution held in Portland, Maine (USA), August 23 and 24, 2002. We were contacted by Irwin SILVERMAN on behalf of EVOLUTION & COGNITION's co-editor, Manfred WIMMER, to have the proceedings published in this journal. The fit seemed most appropriate. Not only are EVOLUTION & COGNITION and NEI founded on an integrative biological evolutionary and cognitive science framework, but on a strongly interdisciplinary perspective as well.

NEI's conception was based on a commitment to bring together divergent viewpoints as, indeed, are reflected in the papers from the conference. The papers is this volume reflect an array of conceptions of unconscious mental states and processes deriving from neuroscience, experimental psychology, philosophy, artificial intelligence, evolutionary biology, and psychodynamic theory. They also reflect the whole rage of investigative norms appropriate to these disciplines.

While many of the papers deal with multiple aspects of unconscious cognition and evolution, we have categorized them by what we think is their main theme. We hope that these papers will not only provide food for thought, but will also convey something of the flavor and excitement of the conference itself.

Paul Bach-y-Rita

Volume Transmission and Brain Plasticity

Abstract

For a century following BROCA's pioneering study on functional localization in the brain, and with the later morphological and physiological studies of CAJAL and SHERRINGTON, the neurosciences were dominated by concepts of synaptic neurotransmission. It was heretical to consider that communication in the brain could be by diffusion. But evidence has been accumulating that a considerable portion of neurocommunication is by diffusion of neuroactive substances through the extracellular fluid of the brain. This has generally been called Volume Transmission (VT).
VT plays a role in space and brain energetics, in drug actions, in recovery from brain damage. VT may be the primary information transmission mechanism in certain normal mass, sustained functions, such as sleep, vigilance, hunger, brain tone and mood and certain responses to sensory stimuli, as well as several abnormal functions, such as mood disorders, spinal shock, spasticity, shoulder-hand and autonomic dysreflexia syndromes, and drug addiction. VT may play a role in the evolution of species.
The difficulties in introducing VT into the conceptual substance of the neurosciences illustrates the effects of the total dominance of a field by a particular concept (in this case the synaptic model). It demonstrates "fashion" aspects of neuroscience, which delays the introduction of radically new concepts (even when strongly supported by experimental evidence) into the Conceptual Substance of the neurosciences. This paper includes material that has been published elsewhere.

Key words

Brain plasticity, volume transmission, brain energetics, brain evolution, sensory substitution, neurotransmission.

Conceptual Issues

Most of the material in this section has appeared in my 1995 book (BACH-Y-RITA 1995). However, since it is obscure and rarely consulted, I consider it appropriate to include it here.

The connectionist-localizationist theory of brain function, based on the neuron doctrine, was the product of many brilliant studies by BROCA (1861); CAJAL (1989); SHERRINGTON (1947), and other scientists of the last half of the 19th and the first third of the 20th century. Anatomical, physiologic, neurochemical and clinical data collected during and shortly after that period fit comfortably within that concept of how the brain works.

However, that model was developed and established before the era of molecular biology; before extrasynaptic diffusion neurotransmission, before G-proteins and ionic and metabotropic membrane signaling systems, before co-transmitters, before glial involvement in neurotransmission, and before the reappearance of brain plasticity concepts after a century of domination of the neurosciences by localizationism. As each of these, and other mechanisms, were demonstrated over the last third of a century, their roles have been interpreted within the existing conceptual substance. The fit was not always comfortable (cf. BACH-Y-RITA 1988, 1995; SHEPHERD 1991), but since there really was no alternative viable theory, it has survived.

An important aspect of this conceptual framework has been the domination of the synaptic model of information transmission. However, synaptic transmission may not be quantitatively the principal means of neurotransmission in the brain: HERKENHAM (1987) considers that receptor-transmitter release site mismatches (in comparison to synapses, at which they are in close opposition) are the rule rather than the exception. Even before the era of molecular biology, the exclusivity of the synapse as a means of transmitting information had been questioned, when results obtained from intra- and extracellular microelectrode studies of polysensory brain stem neurons could not be fit into the prevailing connectionistic theory of brain function (BACH-Y-RITA 1964).

FREEMAN (1992a) considers that the common view of brain organization, as consisting of neuronal networks, has been influenced by the usual GOLGI study illustrations of brain structure. He noted that the GOLGI method stains only about 1.5% of the cells, thus giving an erroneous impression of a discrete network of neurons instead of a densely packed continuous tissue of axons, dendrites and synapses forming a neuropil. Neuronal populations do not exist merely because of large numbers of neurons driven by common input in parallel. They arise by virtue of feedback interactions among cortical neurons that yield cooperative activity. Similarly, GILBERT/HIRSCH/WIESEL (1990) consider that the evidence from earlier GOLGI studies led to the idea that cortical connections were primarily vertical, running across the cortical layers and with relatively little lateral spread. Their studies showed that horizontal processes of pyramidal cells (which allow integration over a large area of the visual cortex) have sparse connections between individual cells; they operate by "mass action". Rather than thinking of receptive fields as being restricted in their extent, with the process of integration of the components of an image occurring at a much later stage along the visual pathway, they have shown that the integrative process is a progressive one, beginning in the primary visual cortex (or perhaps even earlier) and building up in a cascade of diverging and converging connections.

FREEMAN (1984) suggests that perceptual as distinct from sensory information is not expressed at the neuronal level, but rather in the neuronal modules, which consist of assemblies of tens of thousands to hundreds of millions of neurons (FREEMAN 1984). Extensive studies (FREEMAN 1984, 1990, 1992a,b), primarily on the olfactory system as a model of brain function and largely in regard to mechanisms producing EEG signals, has led him to also consider a role for VT in interactions between the cells in a neuron assembly and in synaptically mediated functions, such as the action of norepinephrine in enabling synaptic changes during learning in populations of neurons, by nonsynaptic diffusive release from axons en passant. He views synaptic function as "…the light play of high notes in an organ recitation against the sustained structure of base notes (NDN), both carrying the activity like two legs in walking." (FREEMAN, personal communication). Our studies (BACH-Y-RITA 1991, 1995; BACH-Y-RITA/SMITH 1993) suggest that FREEMAN's "sustained structure of the base notes" may be sufficient in those cases where sustained rather than immediate action is required.

During the century of virtually total domination of the neurosciences by localizationist concepts that followed BROCA's demonstration of functional localization, some scientists presented evidence for plasticity, but their findings were generally ignored. Thus, although CHOW/STEWART (1972) added visual training (rehabilitation) to the HUBEL and WIESEL methods and obtained morphological, physiological and behavioral evidence for recovery, which demonstrated the plasticity of the brain, their study is rarely cited. Our early brain plasticity studies demonstrating the capacity of congenitally blind persons to be trained to perceive visual information delivered to the skin via a modified television camera and an array of vibro- or electrotactile stimulators (BACH-Y-RITA et al. 1969; BACH-Y-RITA 1972, 1989) also were incongruent with the prevailing localizationist concepts[1]. Thus, the prevailing conceptual framework affects the generation of knowledge at all stages, including the choice of experiment, the experimental design, the experimental methods, and the interpretation of the results. In addition, it affects the interest generated by the results, as evidenced by the minimal interest in the CHOW/STEWARD (1972) and other brain plasticity publications.

The prevailing conceptual framework also affects the therapy for specific disease entities: the conclusions of HUBEL/WIESEL (1970) regarding the permanence of amblyopia following lid suture were not likely to have inspired plasticity-based rehabilitation approaches to brain dysfunction due to developmentally-induced abnormalities. Similarly, although FRANZ and associates demonstrated recovery from brain damage with appropriate rehabilitation in both animals and humans and published their results in the Journal of the American Medical Association in 1915 (in which they suggested that we should not speak of permanent paralysis in hemiplegic patients, but rather of untreated paralysis), their study did not influence the clinical management of stroke patients. Furthermore, cerebral ablation experiments leading to conclusions of permanent loss of function have contributed to the long-dominant clinical concept of permanent loss of function, following stroke or traumatic brain damage; little recovery was expected within that conceptual framework, and (consistent with MERTON's (1968) ideas on the "self-fulfilling prophesy") little recovery was obtained (BACH-Y-RITA 1995; BACH-Y-RITA/WICAB-BACH-Y-RITA 1990a,b; COLOTLA/BACH-Y-RITA 2002).

Within that conceptual framework, even most practitioners in the field of Neurological Rehabilitation have sought functional adaptation to the disability resulting from a brain lesion, rather than reorganization of function. Yet in his 1946 book, LURIA (1963) noted that automatic and instinctive reorganizations after a lesion simply exclude the dysfunctional part of the body, while with appropriate rehabilitation, including conscious effort, function is reorganized to include the motorically dysfunctional parts of the body, resulting (within certain limits) in functional recovery.

The concept of a malleable brain, able to recover from damage, was well-demonstrated more than 150 years ago (FLOURENS 1842; he also may have been the first neuroscientist to demonstrate the coexistence of cerebral localization and brain plasticity), and the concept of brain plasticity, although virtually absent from Western science and medicine for more than 100 years after BROCA's (1861) seminal demonstration of cerebral localization (it's effect on the conceptual substance of the neurosciences was discussed elsewhere (BACH-Y-RITA 1990), has been well-accepted in Russia for many years. Pavlov noted in 1932 (cf. LURIA 1963) that brain plasticity has not been accorded its merited importance in nervous system physiology, and LURIA (1946, translated into English in 1963; LURIA et al. 1969) noted that high plasticity is characteristic of functional systems in man, and is the basis for recovery of function following damage. These concepts are now becoming more influential in the clinical management of brain damaged persons.

In this vein, many years ago I stated, in regard to sensory substitution, "The theoretical basis for the design of the vision substitution system… implies a concept of a brain so malleable that the subjective experience of "vision" (as well as the qualitative and quantitative afferent information necessary for useful "vision") could be obtained through an artificial receptor projecting to the cutaneous receptors. Subjective experiences may be the products of a learning process in which afferent inputs from multiple sources are utilized… In the case of information originating from an artificial receptor, the ability of the brain to extract the data and recreate subjectively the image captured by the artificial receptor would depend on the adaptability, or plasticity, of the brain. There is ample evidence… that the brain is capable of adapting to a variety of extreme conditions. That a successful sensory substitution system is not presently in use may not be due to limited functional capabilities of the brain; it may be due to the fact that an artificial receptor system has not yet been constructed to challenge the adaptive capacities of the human brain" (BACH-Y-RITA 1967, p424). Subsequent tactile vision substitution studies have supported those comments (summarized in BACH-Y-RITA et al. 1969; BACH-Y-RITA 1972, 1989).

The re-emergence, during the last few decades, of plasticity concepts in the conceptual substance of the neurosciences has led to essentially co-equal status with localization. Both are essential characteristics of the brain.

It appears that the neurosciences are slower to shift paradigms than other fields. HORROBIN (1990) suggested that this may be due to the absence of an appreciation for theory and hypotheses. He noted that the concept of the dependence of observation on theory and hypothesis is thoroughly understood in physics, chemistry and related sciences; in these fields there is a balanced tension between hypothesis and theory on one hand and experiment on the other. It is recognized that discovery in science almost always begins either with a hypothesis or with an observation that can be seen as anomalous against the background of a clearly defined theoretical construct. HORROBIN noted that the situation in the biomedical sciences could hardly be more different, since we treat ideas in the absence of observation as something that must not be allowed to see the light of day. He further noted that we have no tradition of analytic criticism of existing theoretical concepts, and as a result, general beliefs persist within the biomedical community long after evidence is available to destroy them.

In the absence of broad-based support for theoretical studies within the disciplines of the neurosciences, theories have often been imposed by the strength of character and the professional reputations of the proponents, instead of resulting from open debate in the neuroscience literature. This in no way detracts from the great scientific contributions of those scientists; rather, it reflects the virtual deification of those persons by their followers. Thus, in addition to connectionism and localizationism discussed above, other exclusive theories have been difficult to overturn. One of these is the SHERRINGTONIAN reflex-based concept of cortical function. When experimental evidence from sensory denervated primates demonstrated that motor control could be organized and initiated without sensory input, TAUB (1976) concluded that the reflex-based concept of cerebral function could no longer be supported. However, he reports the anger and rejection with which he was confronted due to this challenge to the hallowed SHERRINGTONIAN dogma (personal communication), although there was no claim that reflex-based and internally-initiated actions are incompatible alternatives.

This example has direct relevance to rehabilitation: TAUB et al. (1994) noted that MOTT and SHERRINGTON in 1895 described the elimination of spinal reflexes by deafferentation (section of the dorsal roots) in monkeys. These studies lead to the position (that became dominant for more than 70 years) that spinal reflexes are the main building blocks out of which behavior is constructed, and that spinal reflexes are necessary for the performance of spontaneous or operant behavior. This kind of negative evidence- that removal or destruction of some part of the nervous system leads to the permanent loss of some functions- completely ignores brain plasticity and the capacity for reorganization. Extensive cortical ablation studies have been interpreted in this negative context, although one of the earliest cortical ablation studies, completed more that 150 years ago, was highly supportive of brain plasticity (FLOURENS 1842). In the case of loss of sensory input, GLANVILLE (1994) considered that the SHERRINGTON concept of neurophysiology no longer explains observed facts. He gave as an example the recovery of a patient with complete absence of proprioception and patchy loss of cutaneous sensation who eventually became totally independent and returned to gainful employment, although at that point routine neurological examination revealed no objective deviation from the pre-recovery state.

Theory that is congruent with modern neurobiology will influence our understanding of mechanisms such as learning, plasticity, drug actions, recovery from brain damage, and of mood, sleep, hunger and other mass sustained functions (BACH-Y-RITA 1991), as well as influencing developments in the fields of artificial intelligence and neuronal networks (AIELLO/BACH-Y-RITA 2001), which are connectivity-based at present. However, the intrinsic limitations in the evolotion of scientific thought will limit the speed at which radically new but experimentally supported concepts enter into the Conceptual Substance of the neurosciences. We scientists are convinced that we are scientific. But fashion is as dominant in science as it is in the field of ladies clothes: when the hemline goes up, everyone follows suit. Similarly, when a concept such as the synapse is firmly established, there is pressure to conform to the accepted dogma.

Volume Transmission

Volume transmission (VT) has also been called nonsynaptic diffusion neurotransmission (NDN). It includes the diffusion through the extracellular fluid of neurotransmitters released at points that may be remote from the target cells, with the resulting activation of extrasynaptic receptors (and possible intrasynaptic receptors reached by diffusion into the synaptic cleft). In contrast to the one-to-one, point-to-point "private" intercelleular synaptic communication in the brain, VT is a slow, one-to-many, widespread intercellular communication (ZOLI et al. 1999). VT may play many roles in the brain in normal and abnormal activity, brain plasticity and drug actions. VT mechanisms may predominate in some functions, probably including the mechanism of action of psychotropic drugs (BACH-Y-RITA 1994). Combinations of both synaptic and diffusion may be the general rule.

VT can be modeled by students in a university classroom (BACH-Y-RITA/AIELLO 1996), who can be equated to neurotransmitter molecules in a vesicle. Upon release, they must go to specific other classrooms spread throughout the campus (receptor sites). They flow out into the halls and the grounds between buildings (extracellular fluid), where they mix with other students (neurotransmitter molecules) from other classrooms (vesicles) going to other target classrooms. They walk (diffuse) to their specific classrooms (receptors) which they enter (bind). In contrast, synaptic communication would require each student to travel to the next class on a motorized separate pathway, much like the enclosed people-transporter bands at the DE GAULLE airport in Paris.

VT appears to be the principal means of neurocommunication in certain mammalian brain regions such the greater limbic system (NIEUWENHUYS 1985) and may play an important role in the organization and regulation of behavior by the core and paracore regions of the brain (NIEUWENHUYS 2000). VT may also be an important mechanism in the highest levels of the human brain (see below). Music appreciation and intellectual functions, for example, may not require the high-frequency (and energetically costly) alternating cycles of activation-inactivation of synaptic transmission, and may be largely replaced by VT (BACH-Y-RITA/AIELLO 2001). VT may play a role in recovery from brain damage, which causes changes in neurotransmitter levels. Some neurotransmitter systems are up-regulated, while others are down-regulated (WESTERBERG et al. 1989). Drug therapy and rehabilitation may induce functional recovery by influencing the affected neurotransmitter systems.

The History Of Volume Transmission

With the development of the techniques of molecular neurobiology, VT has been much more elegantly demonstrated, such as with the immuno-histochemical studies of Fuxe and his collaborators (reviewed by FUXE/AGNATI 1991 and by AGNATI et al. 2000). BEAUDET/DESCARRIES (1978) showed that the biogenic amines released from non-synaptic varicosities may act not only upon adjacent post-synaptic surfaces, but also in tissue of more distant receptor elements. Neurochemical transmission by extrasynaptic diffusion had also been noted by VIZI, who's studies on the non-synaptic modulation of transmitter release (VIVI 1979, 1980, 2000) led him to propose (VIZI 1991) that the essence of brain function (e.g., learning, thinking) lies not in variations of neuronal circuitry (hardware) but rather within the chemical communication itself which is partly wired (synaptic) and partly unwired (nonsynaptic) transmission. I have reviewed these studies elsewhere (BACH-Y-RITA 1993, 1995, 2000, 2002). In the 50s, while studying the action of GABA analogues on thalamic mechanisms in the KILLAM lab at UCLA, we injected them into cat ventricles, since they did not cross the blood-brain barrier (BACH-Y-RITA/HANCE/WINTERS 1960; HANCE et al. 1963). We did not consider diffusion neurotransmission as a mechanism of action, but my later intra- and extracellular microelectrode studies in cat brainstem led to that suggestion. Following sensory stimulation, highly convergent (responding to stimulation of several modalities from various parts of the body) brain cells revealed very long-latency responses, of up to 4 seconds and more. I could not fit the results into classical synaptic mechanism concepts in spite of many surgical and physiological experimental efforts to do so. I suggested that diffusion neurotransmission was the mechanism, and implied that it could play a role in the multiplexing of the polysensory cells (BACH-Y-RITA 1964). This led to a proposed law of the conservation of space and energy in the brain (BACH-Y-RITA 1996).

Brain Mass And Brain Energetics

Assemblies of cells, or neuronal *modules*, have been postulated to form the basis of many functions of the brain (HEBB 1949; SHOLL 1956; FREEMAN 1995; EDELMAN 1992). The cells are separated by a significant volume of extracellular space (NICHOLSON/PHILLIPS 1981). HEBB, (1949 considered that each cell in the assembly is connected to other cells synaptically), by the action of neurotransmitters diffusing through the extracellular space, or by a combination of both. Cell-assembly connectivity in these modules is likely to be varying combinations of synaptic and non-synaptic mechanisms, depending on the specific function.

VT may provide a low-cost space and energy alternative to synaptic communication in the brain. The cost of an action potential has been calculated (AIELLO/BACH-Y-RITA 2000) to be about 10^{11}–10^{12} ATP per cm^2 of depolarized membrane, with a minimum cost of 10^6 ATP at a node of RANVIER. Fully synaptically connected cell-assemblies—of the type HEBB (1949) envisioned to explain mental states, such as thought, expectancy, interest and attention—might exceed metabolic resources: a small (10,000 neuron) module would require approximately 1 joule per liter of brain. Furthermore, the volume of the nerve fibers necessary to fully connect the brain synaptically would create a brain so large that it would not fit in the head (BACH-Y-RITA/AIELLO 1996)

However, rather than purely synaptic or purely diffusive, it is likely that hybrid neuronal networks, exhibiting varying combinations of hard-wired (synaptic) and wireless (VT) connectivities, would fit into the available volume in the brain, with a significant cut in space and energetic costs resulting from removing a significant number of "hard" links (and the nerve fibers needed to connect them) and replacing them with "soft", diffusive pathways. HEBB (1949) noted that integration is fundamentally a question of timing, which has its effect in the functioning of the cell assembly and the interrelation of assemblies. He considered them to be diffuse, anatomically irregular structures that function briefly as closed systems, and do so only by virtue of the time relations in the firing of constituent cells. He could not explain the up to half second delays, which, he noted, is characteristic of thought. However, delays are a characteristic of VT model of cell-assemblies; as noted elsewhere, the initial VT studies revealed delays of up to 4 seconds, which were considered to be related to multiplexing of polysensory pontine cells that could have the effect of reducing the numbers of cells needed for sensory messages (BACH-Y-RITA 1964).

In HEBB's view (1949), waves of depolarization spread along an impressive network of links, carrying information all over the assembly. HEBB did not consider how much energy would be necessary to drive such a network. The estimate depends on the nature of the links. Unmyelinated links require depolarization of the entire link, while myelinated nerves require only depolarization of the short unmyelinated segments at the nodes of RANVIER. The hard-wired scheme easily conflicts with brain metabolic resources. Brain functions requiring the participation of large modules and profuse branching would be incompatible with brain metabolic resources. In this case, additional mechanisms such as VT offer less expensive alternatives to synaptic communication.

Volume Transmission and the Evolution of Species

Physical Anthropologists considering the possibility of energetic constraints on evolution, and exploring where the energy comes from to fuel the large human brain (see Science Research News article entitled "Solving the brain's energy crisis" by GIBBONS 1998) are puzzled by the fact that the human brain does not use more energy than the smaller brains of animals of comparable corporal weight. It is likely that the parts of the human brain that show the greatest size increase over other animals, such as pre-frontal cortex, may be exactly those parts in which highly non synaptic-based functions have their neuronal representation (BACH-Y-RITA/AIELLO 2001). For those functions, such as music appreciation, space-and-energy expensive synaptic neurotransmission may be largely replaced by VT.

The sensory input and the motor output components of human activities such as playing a piano concerto are probably highly synaptically organized (although functions such as vision also have many NDN-mediated mechanisms; BACH-Y-RITA 1995) and do not differ greatly from comparable functions in non-humans. However, components of that activity (playing a piano concerto) are specifically human, such as the musicality and artistic components. These probably involve the specifically human isocortical brain structures, and may not require the high-frequency (and energetically costly) alternating activation–inactivation of synaptic transmission. Although direct evidence is lacking, NDN is consistent with their modes of action.

Falk (in GIBBONS 1998) considered that we have to attend to the energetics or we're not going to get selection for a bigger brain. We suggest that NDN may play a role in evolution, providing a mechanism to allow the underlying physical constraints to be overcome to build an oversize brain (BACH-Y-RITA/AIELLO 2001). MITCHISON (1992) suggested that connectivity appears to be minimized in the brain, and LAUGHLIN et al. (1998) noted that neurons, neural codes and neural circuits have evolved to reduce metabolic demands. Functions that are highly NDN-mediated may be a basis for the reduced per kilogram energy requirements of human brains in comparison to the brains of animals of comparable size.

Conclusion

This review has explored the background of the development of the concept of volume transmission in the brain. It has placed that concept in the context of the evolution of scientific thought, brain mass and brain energetics considerations, brain plasticity and its applications to recovery from brain damage and sensory substitution, and the brain constraints in the evolution of species.

Author's address

Paul Bach-y-Rita, Departments of Orthopedics and Rehabilitation Medicine, and Biomedical Engineering, University of Wisconsin, Madison, WI 53706, USA.
Email: pbachyri@facstaff.wisc.edu

Note

1 Blind persons were able to use visual means of analysis in 2- and 3 dimensional space. Visual illusions and defense reactions to the correctly located spatial information were obtained. Thus, the brain was shown to be capable of learning to mediate "visual" information arriving at the somatosensory cortex, which apparently does not have the specialized image analysis mechanisms of the visual cortex. Whether or not the percepts can be called vision is open to debate. We entitled an early report "Seeing with the skin" (WHITE et al. 1970), and BACH-Y-RITA (1972) and noted that the visual image does not get beyond the retina, from where it is sent to the cortex as pulse-coded neuronal discharges. The individual pulses differ little from any other pulses transmitted along nerves from any region. He questioned whether eyes were necessary for vision. Defining "vision" broadly, he stated "If a subject without functioning eyes can perceive detailed information in space, correctly localize it subjectively, and respond to it … the term 'vision' (can be applied)" However, MORGAN (1977) considered that the blind subjects using a tactile vision substitution system are seeing, not merely "seeing". In any case, considerations such as low resolution, difficulty in perceiving in the presence of complex visual environments, tactile masking, and others, limit the practical applications of tactile vision substitution.

References

Agnati, L./Fuxe, K./Nicholson, C./Sykova, E. (eds) (2000) Volume transmission revisited. Volume progress in brain research. Elsevier Science: Amsterdam.

Aiello, G. L./Bach-y-Rita, P. (2000) The cost of an action potential. Journal of Neuroscience Methods 103:145–149.

Aiello, G. L./Bach-y-Rita, P. (2001) Hebbian brain cell-assemblies: Nonsynaptic neurotransmission, space and energy considerations. In: Cihan, H. D./Buczak, A. L./Embrechts, M. J./Ersoy, O./Ghosh, J./Kercel, S. (eds) Intelligent engineering systems through artificial neural networks. Vol. 11. ASME Press: New York, pp. 441–447.

Bach-y-Rita, P. (1964) Convergent and long latency unit responses in the reticular formation of the cat. Experimental Neurology 9:327–344.

Bach-y-Rita, P. (1967) Sensory plasticity: Applications to a vision substitution system. Acta Neurologica Scandinavica 43:417–426.

Bach-y-Rita, P. (1972) Brain mechanisms in sensory substitution. Academic Press: New York.

Bach-y-Rita, P. (1988) Brain plasticity. In: Goodgold, J. (ed) Rehabilitation medicine. C.V. Mosby Co: St. Louis, pp. 113–118.

Bach-y-Rita, P. (1989) Physiological considerations in sensory enhancement and substitution. Europa Medica Physica 25:107–128.

Bach-y-Rita, P. (1990) Paul Broca: Aphasia and cerebral localization. Current Contents, Life Sciences 33(39):18.

Bach-y-Rita, P. (1991) Thoughts on the role of volume transmission in normal and abnormal mass sustained functions. In: Fuxe, K./Agnati, L. (eds) Volume transmission in the brain. Raven Press: New York, pp. 489–496.

Bach-y-Rita, P. (1993) Nonsynaptic diffusion neurotransmission (NDN) in the brain. Neurochemistry International 23:297–318.

Bach-y-Rita, P. (1994) Psychopharmacologic drugs: Mechanisms of action (Letter). Science 264:642–624.

Bach-y-Rita, P. (1995) Nonsynaptic diffusion neurotransmission and late brain reorganization. Demos-Vermande: New York.

Bach-y-Rita, P. (1996) Conservation of space and energy in the brain. Restorative Neurology and Neuroscience 10:1–3.

Bach-y-Rita, P. (2000) Conceptual issues relevant to present and future neurologic rehabilitation. In: Levin, H./Grafman, J. (eds) Neuroplasticity and reorganization of function after brain injury. Oxford University Press: New York, pp. 357–379.

Bach-y-Rita, P. (2002) Brain damage, recovery from. In: Ramachandran, V. S. (ed) Encyclopedia of the human brain. Vol. 1. Academic Press: New York, pp. 481–491.

Bach-y-Rita, P./Aiello G. L. (1996) Nerve length and volume in synaptic versus difussion neurotransmission: a model. NeuroReport 7:1502–1504.

Bach-y-Rita, P./Aiello G. L. (2001) Brain energetics and evolution. Brain and Behavioral Sciences 24:280.

Bach-y-Rita, P./Collins, C. C./Saunders, F./White, B./Scadden, L. (1969) Vision substitution by tactile image projection. Nature 221:963–964.

Bach-y-Rita, P./Hance A. J./Winters W. D. (1960) Some effects of atropine and eserine on thalamocortical pathways. Proceedings of the Western Pharmacology Society 3:69–79.

Bach-y-Rita, P./Smith, C. U. M. (1993) Comparative efficiency of volume and synaptic transmission in the coerulean system: relevance to neurologic rehabilitation. Scandinavian Journal of Rehabilitation Medicine 25:3–6.

Bach-y-Rita, P./Wicab-Bach-y-Rita, E. (1990a) Biological and psychosocial factors in recovery from brain damage in humans. Canadian Journal of Psychology 44:148–165.

Bach-y-Rita, P./Wicab-Bach-y-Rita, E. (1990b) Hope and active patient participation in the rehabilition environment. Archives of Physical Medicine and Rehabilitation 71:1084–1085.

Broca, P. (1861) Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie. Bulletins et Mémoires de la Société Anatomique de Paris 36:330–357.

Cajal, S. Ramon y (1989) Recollections of My Life. MIT Press: Cambridge.

Chow, K. L./Stewart D. L. (1972) Reversal of structural and functional effects of long-term visual deprivation. Experimental Neurology 34:409–33.

Colotla, V. A./Bach-y-Rita P. (2002) Shepherd Ivory Franz: His contributions to neuropsychology and rehabilitation. Cognitive, Affective & Behavioral Neuroscience 2:141–148.

Edelman, G. M. (1992) Bright Air, Brilliant Fire. Basic Books: New York.

Flourens, P. (1842) Recherches expérimentales sur les propriétés et les fonctions du système nerveux, dans les animaux vertébrés (2nd ed.) Ballière: Paris.

Freeman, W. J. (1984) Premises in neurophysiological studies of learning. In: Lynch, G./McGaugh, J. L./Weinberger, N. M. (eds) Neurobiology of learning and memory. Guilford Press: New York, pp. 231–234.

Freeman, W. J. (1992a) Tutorial on neurobiology: From single neurons to brain chaos. International Journal Bifurcation and Chaos 2:451–482.

Freeman, W. J. (1992b) Predictions on neocortical dynamics derived from studies in paleocortex. In: Basar, E./Bulloch, T. H. (eds) Induced rhythms in the brain. Birkhäuser: Boston, pp. 183–199.

Freeman, W. J. (1995) Societies of brains. Lawrence Erlbaum: New York.

Fuxe, K./Agnati, L. F. (1991) Volume transmission in the Brain. Raven Press: New York.

Gibbons, A. (1998) Solving the brain's energy crisis. Science 280:1345–1347.

Gilbert, C. D./Hirsch, J. A./Wiesel, T. N. (1990) Lateral interactions in the visual cortex. Cold Spring Harbor Symposia on Quantitative Biology 55:663–677.

Glanville, H. J. (1994) What is rehabilitation? In: Illis, L. S. (ed) Neurological rehabilitation (2nd ed). Blackwell: Oxford, pp. 7–13.

Hance, A. J./Winters W. D./Bach-y-Rita, P./Killam, K. F. (1963) A neuropharmacological study of gamma-aminobutyrylcholine, gamma-aminobutyric acid, phyostigmine and atropine. Journal of Pharmacology and Experimental Therapeutics 140:385–395.

Hebb, D. O. (1949) The organization of behavior. John Wiley & Sons: New York.

Herkenham, M. (1987) Mismatches between neurotransmitter and receptor localizations in brain: observations and implications. Neuroscience 23:1–38.

Horrobin, D. (1990) Discouraging hypotheses slows progress. The Scientist, Nov. 26:13–14.

Hubel, D. H./Wiesel, T. N. (1970) The period of susceptibility to the physiological effects of unilateral eye closure in kittens. Journal of Physiology (London) 206:419.

Laughlin, S. B./de Ruyter van Steveninck, R. R./Anderson, J. C. (1998) The metabolic cost of neural information. Nature

Neuroscience 1:36–41.

Luria, A. R. (1963) Restoration of function after brain damage. McMillan: New York. Originally published in 1946.

Luria, A. R./Naydin, V. L./Tsvetskova, L. S./Vinarskaya, E. N. (1969) Restoration of higher cortical function following local brain damage. In: Vinken, P. J./Bruyn, G. W. (eds) Handbook of Clinical Neurology, Vol 3, Disorders of Higher Nervous Activity. North Holland: Amsterdam, pp. 368–433.

Merton, R. K. (1968) The self-fulfilling prophecy. In: Merton, R. K. (ed) Social theory and social structure. Free Press: New York, pp. 475–490.

Mitchison, G. (1992) Axonal trees and cortical architecture. Trends in Neurosciences 15:122–126.

Morgan, M. J. (1977) Molyneux's question. Cambridge University Press: Cambridge.

Nicholson, C./Phillips, J. M. (1981) Ion diffusion modified by tortuosity and volume fraction in the extracellular microenvironment of the rat cerebellum. Journal of Physiology 321:225–257.

Nieuwenhuys, R. (1985) Chemoarchitecture of the brain. Berlin: Springer: Berlin.

Nieuwenhuys, R. (2000) Comparative aspects of volume transmission, with sidelight on other forms of intercellular communication. Volume transmission revisited 125:49–126.

Shepherd, G. M. (1991) Foundations of the neuron doctrine. Oxford University Press: Oxford.

Sherrington C. (1947) The integrative action of the nervous system. Yale University Press: New Haven.

Sholl, D. A. (1956) The organization of the cerebral cortex. Methuen & Co: London.

Taub, E. (1976) Motor behavior following deafferentiation in the developing and motorically mature monkey. In: Herman, R./Grillner, S./Ralston, H. J./Stein, P. S. D./Stuart, D. (eds) Neural control of locomotion. Plenum: New York, pp. 675–705.

Taub, E./Crago, J. E./Burgio, L. D./Groomes, T. E./Cook, E. W./DeLuca, S. E./Miller, N. E. (1994) An operant approach to rehabilitation medicine: Overcoming learned nonuse by shaping. Journal of Experimental Analysis of Behavior 61(2):281–293.

Vizi E. S. (1979) Presynaptic modulation of neurochemical transmission. Progress in Neurobiology 12:181–290.

Vizi E. S. (1980) Nonsynaptic modulation of transmitter release: Pharmacological implications. Trends in Pharmacological Sciences 1:172–175.

Vizi E. S. (1991) Nonsynaptic inhibitory signal transmission between axon terminals: physiological and pharmacological evidence. In: Fuxe, K./Agnati, L. F. (eds) Volume transmission in the brain. Raven Press: New York, pp. 89–96.

Vizi E. S. (2000) Role of high affinity receptors and membrane transporters in nonsynaptic communication and drug action in the central nervous system. Pharmacological Review 52:63–90.

White, B. W./Saunders, F. A./Scadden, L./Bach-y-Rita, P./Collins, C. C. (1970) Seeing with the skin. Perception & Psychophysics 7:23–27.

Zoli, M./Jansson, A./Syková, E./Agnati, L. F./Fuxe, K. (1999) Volume transmission in the CNS and its relevance for neuropsychopharmacology. Trends in Pharmacological Sciences 20:142–50.

Elizabeth Ennen

The Basal Ganglia and the Cognitive Unconscious

Abstract

Neuroscientists have shown that the basal ganglia system in humans supports a form of unconscious "habit" memory. I argue that some forms of human habit memory are both nonrepresentational and cognitive, and thus constitute a challenge to the widely held assumption that the essence of cognition is representational processing. This view has implications for our understanding of the evolution of human cognition, for there are good reasons for thinking that nonrepresentational cognition predates representational cognition.

Key words

Basal ganglia, habit memory, representational theory of the mind, evolution.

HABIT MEMORY IS THE capacity to acquire, retain, and deploy stimulus–response (S–R) associations. Neuroscientists have known for many years that the basal ganglia system in rats and monkeys subserves habit memory. In 1996, Barbara KNOWLTON and her colleagues reported that the basal ganglia system in humans also supports a form of habit memory (KNOWLTON/MANGELS/SQUIRE 1996). The form of habit memory identified by KNOWLTON is now typically described by neuroscientists as *cognitive* in nature. This neuroscientific research is philosophically provocative, for there is a real tension between the claim that habit memory is based on S–R mechanisms and the claim that habit memory is cognitive. The tension arises from the fact that S–R processing is traditionally classified as nonrepresentational, while cognitive processing is traditionally classified as representational. My central claim in this essay is that human habit memory, of the sort described by KNOWLTON, is indeed both cognitive and nonrepresentational and that it thus constitutes a challenge to the widely held assumption that essence of cognition is representational processing. This research can thus be used to support the view, endorsed by a small but growing number of cognitive scientists, that our general theory of the mind should be pluralistic; it should acknowledge both representational and nonrepresentational forms of cognition. Furthermore, this research has implications for our understanding of the *evolution* of human cognition, for, as I will argue, nonrepresentational cognition predates representational cognition in evolutionary time.

The essay will be divided into three sections. In the first, I will provide some background information about the neuroscientific literature on multiple memory systems. Two memory systems are pertinent here: the basal ganglia system, because it is the neural substrate of habit memory, and the hippocampal system, because it serves as an important foil for the basal ganglia system. The first section will therefore be devoted to a general comparison of these two memory systems. In the second section, I will focus specifically on KNOWLTON's research on human habit memory. After describing KNOWLTON's work, I will provide arguments for my claim that human habit memory is both cognitive and nonrepresentational. In the third section, I will discuss the significance of human habit memory for our understanding of the evolution of cognition.

Multiple Memory Systems

During the last fifty years, neuroscientists have identified a handful of distinct neural memory systems. The most well-studied of these systems, the hippocampal system, is said to subserve "declarative" memory—the capacity to acquire, retain, and manipulate information about the world and about one's own life.[1] The hippocampal *system* consists of a neural circuit that involves a number of processing centers, including, most importantly, the hippocampus itself, a small, tubular-shaped structure located in the medial temporal lobe. This system processes both sensory information and propositional information. It thus subserves the capacity to remember, for example, the events of last week, the color and location of a car, the smell of a rose, telephone numbers, and the capitol of France. The consensus in the neuroscientific community is that the hippocampal system accomplishes its work by storing information in distinct neural representations (EICHENBAUM/COHEN 2001). As a result, the hippocampal system is often referred to as the "representational" memory system.

Hippocampal representations exhibit a number of important properties (EICHENBAUM/COHEN 2001). First, they may be acquired very quickly, sometimes after only a single exposure to the relevant "stimulus". So, for example, I can still recall the cover of a book I saw only once last week. Second, hippocampal representations, if sufficiently consolidated, may last for a very long time. I have not seen my childhood copy of *Charlotte's Web* in several decades, but I can still recall both its appearance and its plot. In other words, hippocampal representations may represent an object long after the object has disappeared from the immediate environment. Third, hippocampal representations may be manipulated internally and used in combination to construct "maps" of the external world. Fourth, hippocampal representations are flexible; they may be used repeatedly in many different types of situations. Finally, while hippocampal representations are typically associated with *conscious* mental processes, they may play a role in unconscious processes as well.

The basal ganglia system consists of a number of cortical "loops", all of which pass through the basal ganglia, a collection of sub-cortical nuclei located deep within the brain. (Because one subset of these nuclei is collectively referred to as the "neostriatum", this system is sometimes referred to as the "neostriatal system".) The basal ganglia memory system is said to subserve two types of nondeclarative (or "procedural" memory): skill memory and habit memory. Skill memory is the capacity to acquire, retain, and deploy complex perceptual–motor sequences. Examples of skill memory include the capacity to play an arpeggio on the piano, the capacity to juggle, and the capacity to dribble a basketball.[2] Habit memory, in contrast, is the capacity to respond automatically to a particular stimulus, by deploying a response that has been associated with the stimulus in the past. The basic dynamic of habit memory acquisition is the repetitive association of a particular response with a particular stimulus via the technique of reinforcement.

The concept of habit memory is best explained via an example. Consider the following "radial arm maze" experiment (MCDONALD/WHITE 1993).[3] A radial arm maze is a large elevated maze consisting of a round, central platform and a number of "radiating" arms. In this experiment, an eight-arm maze is used; each of the maze arms is equipped with a light at the entrance of the arm and a food well at the end of the arm farthest from the central platform. On the first day of the experiment, four of the eight arms are randomly chosen to be baited with food and lit. (The other four arms remain unbaited and unlit.) A rat is then placed on the maze and allowed to explore. If the rat finds and consumes a reward, the reward is replaced one time. When the rat retrieves the second food reward from a particular maze arm, the food well is left empty and the light is turned off. After the rat has found and consumed the eight food rewards, the rat is removed from the maze.[4] On the second day of the experiment, the scientists again select four arms randomly and bait and light the selected arms. Once again the rat is allowed to explore the maze until it has found and consumed the eight food rewards. This process continues for a number of days.

MCDONALD and WHITE found that normal rats gradually learn to go directly to the lit arms and to avoid the unlit arms. Their foraging thus becomes extremely efficient. Rats with basal ganglia damage, in contrast, do not become more efficient. They continue to explore both lit arms and unlit arms. MCDONALD and WHITE analyze the results of this experiment in terms of habit memory. They note that whenever a rat explores a lit arm, it finds a food reward. Over time, they suggest, a stimulus–response association is gradually formed between the "light" stimulus and the "approach the light" response. After a fair amount of experience with the maze, normal rats use habit memory to forage on the maze efficiently. Basal ganglia rats fail on this task, according to MCDONALD and WHITE, because they are incapable of forming the necessary stimulus–response associations.

Habit memory, as it is manifest in this task, displays a number of critical properties that are quite different from the properties of declarative, or hippocampal, memory. First, while hippocampal memory can be acquired very quickly, basal ganglia memory is always acquired very slowly; it takes a fair amount of time for the association between the stimulus and the response to form. Second, in the case of habit memory, the basal ganglia system does not generate a distinct neural representation. Instead, it works directly on perceptual and motor performance systems, making small, incremental changes gradually over time. These changes affect the way the animals reacts to the stimulus in the future. So while the hippocampal memory involves the creation of a distinct representation, basal ganglia memory results in the creation of a *disposition*. Third, habit memory is inflexible. During the acquisition phase, neural changes are made directly to performance systems while the animal is engaged in a particular type of behavior. These neural changes are manifest only when these performance systems are re-engaged in that same type of behavior. Finally, while hippocampal memory is typically associated with conscious thought, basal ganglia memory is always *deployed unconsciously.* In the radial arm maze experiment, the rat's nervous system is modified in such a way that the appearance of the light at the end of a maze arm triggers the "approach the light" behavior *automatically.* In the case of habit memory, the rat does *not* learn a piece of propositional information to the effect that "where there is light, there is food". The rat simply responds, automatically, to the light stimulus.

Neuroscientists have been studying the relationship between the basal ganglia system and habit memory in rats and monkeys for several decades. Attempts to identify a similar form of habit memory in humans have been stymied by the fact that the hippocampal system is so powerful in humans that it typically steps in and performs "habit" tasks using its capacity to generate propositional representations. If a human subject were asked to participate in the radial arm maze task described above, for example, the subject would be able to generate and retain the proposition "where there is light, there is food" very quickly. As a result, the experiment would not allow scientists to determine whether or not the subject possessed a habit memory system. In 1996, KNOWLTON described the experimental impasse as follows: "a neostriatal habit learning system like the one identified in rodents has not been demonstrated in humans" (KNOWLTON/MANGELS/SQUIRE 1996, p1399).

Habit Memory in Humans

KNOWLTON and her colleagues made a breakthrough by designing an S–R task in which the relationships between the stimuli and the responses are too "complex" for the hippocampal system to process in a declarative fashion (KNOWLTON/MANGELS/SQUIRE 1996; KNOWLTON/SQUIRE/GLUCK 1994).[5] The stimuli, or "cues", in this experiment are a number of rectangular cards. Each card in the deck displays a number of geometrical figures, all of the same sort. So, for example, some of the cards display a number of squares. There are four types of cards in total: the "square" cards, the "triangle" cards, the "diamond" cards, and the "circles" cards. During the experiment, human subjects are seated before a computer monitor. On each trial, one, two, or three cards (each of a different type) are shown on the monitor and the subject is asked to state whether that particular combination of cards is predictive of "sunshine" or "rain". After the subject has made her prediction, a picture of a face appears on the monitor. If the subject has made a correct prediction, a picture of a smiling face appears on the monitor. If she has made an incorrect prediction, a picture of a frowning face appears instead. The task for the subject, then, is to grasp the relationship between the cues (the card combinations) and the two possible outcomes (sunshine and rain). To put the same point in slightly different terms, the subject must learn to make specific responses (e.g. "sunshine") to particular stimuli (various card combinations).

What makes the weather prediction task extremely difficult is the fact that the relationships between the cues and the outcomes are very complicated. First, each type of card is *probabilistically* related to each of the two outcomes. So, for example, square cards are highly predictive of sunshine. (In 75% of the trials in which a square card appears, the correct prediction is "sunshine".) Triangle cards, on the other hand, are only somewhat predictive of sunshine. (In 57% of the trials in which a triangle card appears, the correct response is sunshine.) Diamond cards are highly predictive of rain; circle cards are somewhat predictive of rain. Second, the correct response on any given trial is a function of either one, two, or three card types, depending on how many cards are presented to the subject on that trial. Given these complexities, the relationship between the cues and the outcomes cannot be easily grasped in a declarative fashion. As a result, the task provides an opportunity for the habit system in humans to perform.

When normal subjects participate in this weather prediction task, their performance improves steadily over time. After fifty trials, normal subjects typically respond correctly in 70–75% of the trials, even though they claim, particularly at first, to be guessing. Subjects with hippocampal damage are capable of learning this task, which indicates that successful performance does not depend on the integrity of the hippocampal memory system.[6] Subjects with basal ganglia damage, on the other hand, are severely impaired on this task. KNOWLTON concludes that "in humans, the neostriatum … is essential for the gradual, incremental learning of associations that is characteristic of habit learning" (KNOWLTON/MANGELS/SQUIRE 1996, p1399).[7] She suggests, in other words, that human subjects learn to associate the verbal responses "sunshine" and "rain" with certain card combinations in the same way that rats learn to associate the perceptual–motor response "approach the arm" with the stimulus "lit arm". Just as the rat's behavior is gradually modified via the use of a food reward, the human subject's behavior is gradually modified, on this account, via the use of a visual reward, the smiling face. Just as the rat's behavior is generated "automatically", the human subject's behavior is generated using mechanisms that are not available to consciousness. Subjects do well even when they are just "guessing".

In some contexts, KNOWLTON speaks of the weather prediction task as a "classification" task. The subject is, in a sense, classifying all of the various card combinations into two distinct categories: the sunshine category and the rain category. We can think of habit memory in rats in similar terms, for when the rat is participating in the radial arm maze experiment, it classifies maze arms into two categories: baited arms and unbaited arms. KNOWLTON's colleague, Larry SQUIRE, makes a provocative suggestion based on this sort of analysis. He notes that the human capacity to distinguish "fine wines" from "mediocre wines" (or "authentic paintings" from "forgeries") may be based on the same neural mechanisms as those involved in the rat's capacity to distinguish between baited and unbaited arms (SQUIRE/KANDEL 2000, pp179–180). SQUIRE's musings suggest that habit memory may be involved in very sophisticated forms of behavior.

The neuroscientific research on habit memory deserves careful philosophical scrutiny. My goal here is to provide support for two "philosophical" claims: first, that habit memory is genuinely nonrepresentational, and second, that habit memory is genuinely cognitive.[8] One reason for supposing that habit memory is nonrepresentational is that at least some neuroscientists describe it as such. Mortimer MISHKIN is explicit on this point when he notes that "what is stored in the habit formation system is not the neural representations of such items as objects, places, acts, emotions, and the learned connections between them but simply the changing probability that a given stimulus will evoke a specific response due to the reinforcement contingencies operating at that time" (MISHKIN/MALAMUT/BACHEVALIER 1984, p72). Neuroscientific usage is neither uniform nor philosophically rigorous, however, and so I will need a stronger argument for the claim that habit memory is nonrepresentational. After all, one might argue that habit memory involves the formation of an association between a *representation* of the stimulus and a *representation* of the response. In order to respond to this concern, I will present a principled method for distinguishing between internal states that are representational, and internal states that are nonrepresentational. For help in this regard, I turn to Andy CLARK, a cognitive scientist who has recently offered an account of the conditions under which it is appropriate to call an internal state a *genuine* representation.

On CLARK's view, a genuine representation is an information-bearing internal state that can be decoupled from its source object (i.e. from the object which it is representing) (CLARK 1997a, p463; CLARK/GRUSH 1999, pp7–8). In other words, an information-bearing internal state counts as a genuine representation only if it has the capacity to serve as a surrogate for an object that is not present in the environment. Because such internal states can be decoupled from the world, so to speak, they can be manipulated internally in a flexible way and, as CLARK points out, they may be used in combination to construct maps of the external world. Genuine representations can be contrasted with what I call "faux" representations—information-bearing internal states that fall silent when their source objects are absent from the immediate environment. Consider the following example. If a person were to reach out to grasp a mug of coffee, she would no doubt come to possess internal states that convey information about the mug. Some of these internal states might qualify as genuine representations. But others might not, since some of these information-bearing internal states might be functional only in the presence of the mug. As a result, they would not be capable of serving as surrogates for the mug (or for the mug's salient properties).

Why should we draw the line between representations and nonrepresentations in the way CLARK suggests? CLARK's position is compelling. He notes that cognitive scientists are vulnerable to the charge that, given their theoretical commitments, they are liable to see representations everywhere they look, whether or not there are any representations present. CLARK argues that this vulnerability can be mitigated by reserving the notion of representation for internal states that animals (including humans) *use* as representations, i.e. as surrogates for absent objects.

The distinction between genuine representations and faux representations is germane to the issue of multiple memory systems. The hippocampal system is a genuine representational system *par excellence*. Because hippocampal representations are long-lasting, they can, in effect, be de-coupled from their source objects. Furthermore, as we have already seen, hippocampal representations may be used to construct maps of the external world. The basal ganglia system, on the other hand, does not traffic in genuine representations. Its internal states, while information bearing, are dependent on environment in a way that the information-bearing states of the hippocampal system are not. S–R associations are functional only in the presence of the relevant stimulus. If you were to remove the stimulus from the environment, the neural substrate of the S–R association (i.e. the various modifications made to performance systems) would remain inert. The internal states of the basal ganglia system cannot, therefore, serve as surrogates for objects in the environment. As a result, they do not qualify as genuine representations. Habit memory is thus nonrepresentational.

I now turn to the question of whether or not habit memory in humans is, at least in some instances, cognitive in nature. An interesting philosophical problem arises, however, at this point. For the last fifty years, mainstream cognitive science has been motivated by the assumption that cognition involves the computational manipulation of representations. Many cognitive scientists view representational processing as a necessary condition on cognition. For proponents of this view, the claim that habit memory is nonrepresentational simply *entails* the claim that it is not cognitive. Representationalism of this sort is still popular in cognitive science today, but there is an ever-increasing amount of dissent. Dynamic systems theorists, for example, argue that at least some (if not most) cognitive processing occurs *nonrepresentationally* (CLARK 1997a, p461). The so-called "embodied and embedded" cognitivists, who are influenced by phenomenologists such as HEIDEGGER and MERLEAU-PONTY, argue that cognitive science should call into question its commitment to representationalism and allocate resources to exploring alternative, nonrepresentational views of the mind (CLARK 1997b). In light of this increasing diversity of views, we should reject the claim that all cognition is *necessarily* representational. The following question should then be construed as open and empirical: to what extent does the mind reply on representational processing for its cognitive achievements? If we take this approach, we open the door to the possibility that successful performance on the weather prediction task is both nonrepresentational and cognitive.

Is human habit memory cognitive? This question is complex and I cannot do it full justice here, but I will gesture at three reasons for thinking that skillful performance on the weather prediction task *is* a manifestation of cognitive abilities. First, it is worth noting that it is common neuroscientific practice to refer to the capacity manifest in the weather prediction task as a "cognitive skill". (See, for example, GABRIELI 1998 and KNOWLTON/SQUIRE/GLUCK 1994.) This is due, at least in part, to the fact that the skill transcends the perceptual–motor realm. But it is also due to the fact that successful performance on this task appears to involve intelligent behavior of a relatively sophisticated sort. The fact that the alleged cognition is not consciously accessible should not deter us, given that we often describe the inner processes of computers as "cognitive". Second, as SQUIRE notes, we can imagine the possibility that certain very sophisticated forms of behavior depend on the habit memory mechanisms of the basal ganglia system (SQUIRE/KANDEL 2000, pp179–180). Third, the view that at least some forms of cognition are dispositional (and not representational) is not without precedent in the philosophical literature on the mind. (See, for example, RYLE 1949). The neuroscience literature on multiple memory system suggests that both RYLEAN dispositionalism and contemporary representationalism are correct, as long as each is applied only to certain types of cognition.

I am now in a position to clarify and reformulate my thesis. I have argued that basal ganglia habit memory is both nonrepresentational and cognitive. Because basal ganglia memory in humans is, at least in some instances, cognitive, it is reasonable to refer to the phenomenon of "basal ganglia cognition". If this terminological sleight-of-hand is permitted, the central claim of the essay becomes a slightly more compact assertion: basal ganglia cognition is nonrepresentational. I now turn to the task of examining the implications of this view for our understanding of the evolution of cognition.

Habit Memory and the Evolution of Cognition

In his book, *Implicit Learning and Tacit Knowledge*, Arthur REBER argues that unconscious cognitive processes predate, in evolutionary time, conscious cognitive processes. According to REBER, "unconscious, implicit, covert functions must have antedated conscious functions by a considerable period of time" (REBER 1993, pp88). I will not rehearse REBER's arguments for this claim here. Instead, I will accept his view on the "primacy of the implicit" and offer a complementary principle regarding the relative primacy of the nonrepresentational.

The classic examples of unconscious cognition involve *representational* cognition. REBER's account of implicit learning, for example, focuses on examples in which a subject has learned some sort of "information". Basal ganglia cognition, however, is a form of *nonrepresentational* unconscious cognition. There is no "information" lurking in the system; there is only a disposition to respond to the world differently "next time". Where does *this* type of unconscious cognition fit into the evolutionary story? We know that the neural structures involved in habit memory predate the neural structures involved in declarative memory. REBER, drawing on the work of Larry SQUIRE (1986), makes the point as follows: "(SQUIRE) pointed out that primitive associative processes, which represent the first form of a procedural system, are handled by neuronal systems which occur first in vertebrates, while declarative systems, which involve consciousness and awareness, require the elaboration of the medial temporal structures, including the hippocampus" (REBER 1993, pp80–81). We may thus infer that basal ganglia cognition predates, in evolutionary time, representational cognition.

Neuroscientific work on multiple memory systems may be relevant for our understanding of the evolution of cognition in another context as well. Cognitive scientists are currently involved in a debate over the degree to which cognition is "continuous" with early and basic forms of associative learning. The work on multiple memory systems allows us to generate the following hypothesis: perhaps different types of cognition bear different relationships to early forms of associative learning. It would be sensible to suppose that cognition subserved by the basal ganglia system is indeed continuous with early forms of associative learning. On the other hand, the literature on multiple memory systems offers no support for the view that cognition based on the hippocampal system is continuous with early forms of associative learning.

Conclusion

KNOWLTON's research on human habit memory is scientifically and philosophically important, for it suggests that some forms of human memory are both nonrepresentational and cognitive (KNOWLTON/MANGELS/SQUIRE 1996; KNOWLTON/SQUIRE/GLUCK 1994). Given the current state of flux in cognitive science, this neuroscientific research comes at an ideal time. Neuroscientific accounts of human habit memory provide additional support for the claim that our best account of cognition will be pluralistic; it will acknowledge the role of both representational and nonrepresentational processes in the generation of intelligent thought and behavior. This research can help us distinguish *which* types of cognition are representational and, importantly, which are not. Furthermore, this neuroscientific research may help us draw a more complete picture of the evolution of human cognition.

Author's address

Elizabeth Ennen, Department of Philosophy, Twilight Hall, Middlebury College, Middlebury, VT, 05753, USA.
Email: ennen@middlebury.edu

Notes

1 Scientists use the term "declarative" for this type of memory because humans can typically "declare" or describe the contents of hippocampal representations.

2 Recent research suggests that the basal ganglia subserves a number of "cognitive" skills as well. See, for example, GRAYBIEL (1998) and GRAYBIEL (1997).

3 The actual experiment is complex; it is designed to uncover a triple dissociation among three different memory systems. I describe only one part of the experiment here.

4 If, after ten minutes, the rat has not retrieved the eight food rewards, it is removed from the maze anyway.

5 KNOWLTON and her colleagues first described the weather prediction task, which is based on work conducted by GLUCK/BOWERS (1988), in 1994. In 1996, KNOWLTON and her colleagues reported that the basal ganglia system is critical for normal performance on this task.

6 After fifty trials, control subjects begin to outperform amnesics in a way that suggests that the hippocampal system does eventually contribute to performance on this task.

7 Recent fMRI studies have confirmed that the basal ganglia system is implicated in the weather prediction task (POLDRACK et al. 1999).

8 I am speaking here of the form of habit memory studied by KNOWLTON in the weather prediction tasks. Other types of human habit memory may well be noncognitive.

References

Clark, A. (1997a) The dynamical challenge. Cognitive Science 21(4):461–481.

Clark, A. (1997b) Being there: Putting brain, body, and world together again. MIT Press: Cambridge MA.

Clark, A./Gush, R. (1999) Towards a cognitive robotics. Adaptive Behavior 7(1):5–16.

Eichenbaum, H./Cohen, N. (2001) From conditioning to conscious recollection: Memory systems in the brain. Oxford University Press: Oxford.

Gabrieli, J. (1998) Cognitive neuroscience of human memory. Annual Review of Psychology 49:87–115.

Gluck, M./Bower, G. (1988) From conditioning to category learning: An adaptive network model. Journal of Experimental Psychology General 117(3):227–247.

Graybiel, A.(1997) The basal ganglia and cognitive pattern generators. Schizophrenia Bulletin 23(3):459–469.

Graybiel, A. (1998) The basal ganglia and chunking of action repertoires. Neurobiology of Learning and Memory 70(1–2):119–136.

Knowlton, B./Squire, L./Gluck M. (1994) Probabilistic classification learning in amnesia. Learning and Memory 1:106–120.

Knowlton, B./Mangels, J./Squire, L. (1996) A neostriatal habit learning system in humans. Science 273(5280):1399–1402.

McDonald, R./White, N. (1993) A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. Behavioral Neuroscience 107(1):3–22.

Mishkin, M./Malamut, B./Bachevalier, J. (1984) Memories and habits: Two neural systems. In: Lynch, G./McGaugh, J. L./Weinberger, N. M. (eds) Neurobiology of learning and memory. Guilford: New York, pp. 65–77.

Poldrack, R./Prabhakaran, V./Seger, C./Gabrieli, J. (1999) Striatal activation during acquisition of a cognitive skill. Neuropsychology 13(4):564–74.

Reber, A. (1993) Implicit learning and tacit knowledge: An essay on the cognitive unconscious. Oxford University Press: Oxford.

Ryle, G. (1949) The concept of mind. Penguin: New York.

Squire, L./Kandel, E. (2000) Memory: From mind to molecules. Scientific American Library: New York.

Stephen William Kercel

Endogenous Causes—Bizarre Effects

Abstract

The defining characteristic of complex systems appears to be endogeny, the quality of making themselves up as they go along. The entailment structure of an endogenous system requires a multi-level self-referential system loop of causality including a closed loop of efficient cause. The loop of efficient cause produces internal semantic behavior. The multi-level nature of the system is reflected in the fact that at least one of its causal entailments is entangled, the same event serving simultaneously serving as both material cause of one effect and efficient cause of another.

Internal semantic behavior does not occur in mechanisms or algorithms. Nevertheless, inferential entailment structures that produce it have been known to mathematicians for at least a century. Such structures are called impredicative systems. Impredicative abstract structures provide a rational way to understand endogenous natural processes.

The effects or behaviors of endogenous and impredicative organizations are bizarre, but not absurd. Even the organizational structures themselves are bizarre. They are multi-level closed-loop hierarchies in which each one of the organization's finitely many nodes is exactly in the middle of the structure. Each node corresponds to an infinitely large cluster of entailments. The causal entailment structure of an endogenous natural system is similar to the inferential entailment structure in an impredicative mathematical system.

Key words

Endogenous, impredicative, complex, hyperset, superphysics, abduction.

Introduction

Might it be that the traditional concepts of "hard science" and their computational analogs are inadequate to understand the processes of life and mind? This concern is not new; no less a light than SCHRÖDINGER suspected that it might be the case. "We must be prepared to find a new type of physical law prevailing in it. Or are we to term it a non-physical, not to say a super-physical, law?" (SCHRÖDINGER 1944, p80). In the sixty years since SCHRÖDINGER's pronouncement, the greatest minds in both physical and computational science have searched for a demonstration that neither such a super-physical law nor its mathematical analogs are necessary. After all that time, during which billions of dollars and uncountable intellectual effort have been poured into the process, it is reasonable to ask how the search is faring.

If we are to judge by the words of one of the most distinguished leaders of the search, Dr. Rodney BROOKS, Director of MIT's Artificial Intelligence Laboratory, it is not going very well. "But maybe there's more to us than computation. Maybe there's something beyond computation in the sense that we don't understand and we can't describe what's going on inside living systems using computation only. When we build computational models of living systems—such as a self-evolving system or an artificial immunology system—they're not as robust or rich as real living systems. Maybe we're missing something, but what could that something be? … My working hypothesis is that in our understanding of complexity and of how lots of pieces interact we're stuck at that algebra–geometry stage. There's some other tool—some organizational principle—that we need to understand in order to really describe what's going on" (BROOKS 2002). This is an astounding assessment of progress in the field. The observation that maybe there is something beyond computation, and maybe living processes are governed by some other organizational principle than one that can be accounted for by computation is tantamount to supposing that SCHRÖDINGER guessed right about the need for a super-physical law.

The fact that something beyond computation is operating in brains is clear from the observation of nematode brains. "They have 300 neurons and are very simple. People know exactly the developmental pattern. There is a research group at MIT, which has been trying to figure out why the stupid little worm does the thing that it does. We know entirely about its developmental pattern. We know all of its neurology, but nobody can figure out what the heck it's doing and why. They did try connectionist models and they gave them up quickly because they just abstract too far away from the physical properties of the nervous system" (CHOMSKY 1993, p86).

If this is the case in the simplest of brains, what is observed in the most complicated of brains? The most surprising observation is that they are *not* "neural networks." While it is the case that brain structure includes neurons, and many neurons form a network interconnected by synapses, the characterization that the brain is a "neural network" only begins to describe the operation of brain function. Nonsynaptic diffusion neurotransmission (NDN, also called "volume transmission" in the medical literature) accounts for a substantial fraction of brain activity. "NDN may be the primary information transmission mechanism in certain normal mass, sustained functions, such as sleep, vigilance, hunger, brain tone and mood, and certain responses to sensory stimuli, as well as several abnormal functions, such as mood disorders, spinal shock, spasticity, shoulder-hand and autonomic dysreflexia syndromes, and drug addiction" (BACH-Y-RITA 1995, p21). In other words, some of the brain's more interesting activities are *not* the result of the operation of its "neural network." Furthermore, they have never been characterized mathematically.

In the face of this sort of evidence, one might rather ask why it took sixty years to come back to the supposition that SCHRÖDINGER might have been right to begin with. It is an unfortunate reality that those of us who are in the business of "hard science" are biased by a far older presumption. We labor under the error of CARTESIAN dualism, a 400-year-old tradition that continues to confuse scientific inquiry even to the present day.

DAMASIO illuminates the nature of the confusion. Based on decades of observation of brain function, he draws the following conclusion. "This is DESCARTES' error: the abyssal separation between body and mind, between the sizable, dimensioned, mechanically operated, infinitely divisible body stuff, on the one hand, and the unsizable, undimensioned, un-pushpullable, nondivisible mind stuff; the suggestion that reasoning, and moral judgment, and the suffering that comes from physical pain or emotional upheaval might exist separately from the body. Specifically: the separation of the most refined operations of mind from the structure and operation of a biological organism" (DAMASIO 1994, pp249–250).

Can we recover from DESCARTES' error? Must we regard the processes of life and mind as a computable mechanism inhabited by a magical and ghostly *élan vital*, hopelessly beyond our ability to understand? Is it not more reasonable to suppose that SCHRÖDINGER was right about the need for a superphysical law of life and mind? Is that supposition not made even more reasonable by BROOKS' assessment that the sixty-year search for an explanation for the processes of life and mind within the bounds of computation may have come up empty, and that perhaps the time has come to look beyond computation?

If BROOKS says that we must look beyond computation, and if our previous looking has been limited by a CARTESIAN bias, then DAMASIO's observations of mental function give us a fairly good suggestion for where we should be looking. According to DAMASIO, it is an observed empirical fact that moral judgment, suffering and emotional upheaval are inseparable from the physical substrate that experiences them. These are semantic values entangled in a substrate.

Semantics is very likely a place that VON NEUMANN would have looked, had he not died prematurely. He came to realize that traditional non-semantic mathematical formalisms do not describe brain function. Speaking of brains he says, "There exist here different logical structures from the ones we are ordinarily used to in logics and mathematics" (VON NEUMANN 2000, p82).He suspected that semantics were crucial to living processes. "By axiomatizing automata in this manner, one has thrown half the problem out the window, and it may be the more important half" (VON NEUMANN 1966, p77). *Following VON NEUMANN's lead, to rationally understand SCHRÖDINGER's "super-physical," and BROOKS' "beyond computation" properties of brain/mind function, let us consider its semantics.*

Are There Semantic Logical Structures?

"Terms like 'yes,' 'no,' 'true,' 'false,' 'fact,' 'reality,' 'cause,' 'effect,' 'agreement,' 'disagreement,' 'proposition,' 'number,' 'relation,' 'order,' 'structure,' 'abstraction,' 'characteristic,' 'love,' 'hate,' 'doubt,' etc., are such that if they can be applied to a statement they can also be applied to a statement about the first statement, and so, ultimately, to all statements, no matter what their order of abstraction is. Terms of such a character I call *multiordinal terms.*

The main characteristic of these terms consists of the fact that on different levels of orders of abstractions they may have different meanings, with the result that they have no general meaning; for their meanings are determined solely by the given context, which establishes different orders of abstractions" (KORZYBSKI 1994, p14). In other words, semantics is more than a "this for that" substitution of names. Semantic concepts are context dependent, having different meanings on different levels of abstraction. A sign has one and only one meaning only in the unusual context leading to the degenerate case of non-ambiguity.

VON NEUMANN saw semantics as going "out the window," as soon as ambiguity is disallowed. In fact, it is in computational algorithms that ambiguity is most strictly disallowed. "Each step of an algorithm must be precisely defined; the actions to be carried out must be rigorously and unambiguously specified for each case" (KNUTH 1973, p5).

Does the fact that ambiguity is banned from algorithms mean that it must be banned everywhere? A rational understanding of a coherent ontological process requires non-contradictory epistemological description, but must that non-contradiction be obtained at the price of disallowing ambiguity? Traditionally, reductionism uses unambiguity in the hope of obtaining non-contradiction, but is this strategy reasonable?

The usual epistemological device for non-contradictory representation is a structure technically known in mathematics as a *formalism*. A formalism is an unambiguous list of finitely many true propositions. As a consequence of GÖDEL's theorem, if (as is almost inevitable in computation) these propositions formalize a number theory, then they must refer to at least one external referent (SIPSER 1997, pp209–210). Since the formalism is unambiguous, the propositions may refer to no more than one external referent. Thus, a formalism has a semantic meaning; it is about the external referent.

An internal semantic referent, in addition to the required external referent, is disallowed since it would also render the formalism ambiguous. In set theory, this ban takes the form of the Foundation Axiom, the notion that no set may include itself as a member. (For those who quibble over the generality of GÖDEL's theorem: The axioms of set theory imply the existence of the VON NEUMANN ordinals, which provide a set-theoretic definition of integers. GÖDEL's theorem applies; sets are about something.) The strict insistence on non-ambiguity in a formalism is supposed make it a tool for performing non-contradictory reasoning. "By formalizing the theory, the development of the theory can be reduced to form and rule. There is no longer ambiguity about what constitutes a statement of the theory, or what constitutes proof in the theory" (KLEENE 1950, p63).

The Foundation Axiom is a notable example of imposing a ban on ambiguity in the hope of eliminating contradiction. The notorious Russell Paradox is simply the observation that if there exists an object that constitutes the "set of all the sets," then provably, it both *is* and *is not* a member of itself. Clearly, this is a contradiction, and what it actually proves is that there is no such object as "set of all the sets" in the formalism defined by the set axioms (excluding the Foundation Axiom). In the hope of precluding the possibility of contradiction, the Foundation Axiom imposes a general ban on a class of inherently ambiguous sets; specifically it disallows any set that includes itself as a member. (GOLDREI 1996, pp68–70)

However, the Foundation Axiom is too restrictive. A more general theory of *hypersets* is afforded by the Anti-Foundation Axiom. (For the genuinely curious, the simplest form of the Anti-Foundation Axiom says that "every tagged graph has a unique decoration." The explanation of this cryptic definition is beyond the scope of this paper, but is readily accessible in *The Liar.*) (BARWISE/ETCHEMENDY 1987, pp39–44). Hyperset theory is provably no less coherent than conventional set theory, and sets turn out to be a degenerate case of hypersets.

In set-theoretic jargon, the prototypical hyperset, Ω, is the set that contains itself as a singleton.

More casually, Ω is that object which solves $\Omega = \{\Omega\}$, where the brackets {} denote a set. This definition is no more a conceptual stretch than the definition of i as that object which solves $i^2 = -1$. The uniqueness of Ω is proven; there is one and only one object that satisfies the definition of Ω, and it differs from all other objects in mathematics (ACZEL 1988).

$\Omega = \{\Omega\}$ is not a circular definition, or a claim that a thing merely is what it is. It is an *impredicative* definition. The identity of Ω is established by the specific constraint on its relationship with itself.

The conventional meaning of the term impredicative is given by KLEENE. "When a set M and a particular object m, are so defined that on the one hand m is a member of M, and on the other hand the definition of m depends on M, we say that the procedure (or the definition of m, or the definition of M) is *impredicative*. Similarly, when a property P is possessed by an object m whose definition depends on

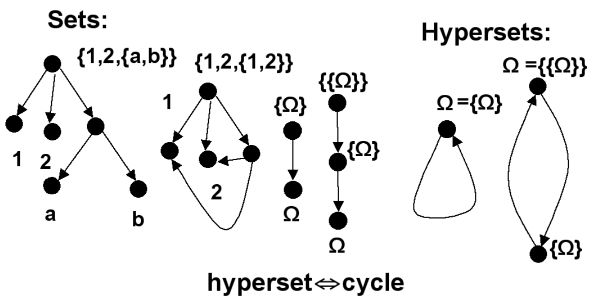

Figure 1. Graphical representation of sets and hypersets.

P (here *M* is the set of the objects which possess the property *P*). An impredicative definition is circular, at least on its face, as what is defined participates in its own definition" (KLEENE 1950, p. 42).

Note that KLEENE does not dismiss impredicativity because it is circular "on its face." It has depth below the face; the object on one level refers to its definition on the other, and there is a specific constraint on how that reference operates across the two levels. The distinguishing feature of an impredicative definition is the multi-level constraint. KLEENE shows that the constraint in the impredicative definition can be used to define the least upper bound on a set of real numbers. Impredicativity is not merely a theoretician's toy; for example, modern telephone engineering is based on Wavelets, and the scaling function in the Fundamental Wavelet Equation is impredicatively defined. (AKANSU/HADDAD 1992, p315)

Impredicativity has been known to mathematicians for over a century, and there are various philosophical objections to the concept. Addressing those objections is beyond the scope of this paper. However, they are readily answerable and have been answered (PICARD 1993). Hypersets, as an instance of impredicative objects, are legitimate and useful mathematical objects.

The easiest way to gain a sense of hypersets is to consider the graphical representation of sets (Figure 1). A set is a sort of abstract container that can define a hierarchy. 1 is a number; it is not a set. {1}is the set whose only member is the number 1. Sets can be nested; {{1}}is the set whose only member is the set whose only member is the number 1.

The hierarchy defined by set nesting is represented by a directed graph. A terminal node in the graph represents a member of a set, and not usually a set itself. A non-terminal node represents a set; directed edges from the node that represents the set are connected to all the nodes that represent members of the set. The leftmost object in Figure 1 is the set containing three members; those members are the numbers 1 and 2 and the two-element set whose members are the letters *a* and *b*. The same members may appear at different levels in the hierarchy. The next leftmost object in Figure 1 is another set containing three members; those members are the numbers 1 and 2 and the two-element set whose members are the numbers 1 and 2.

The two objects in the middle of Figure 1 are not hypersets. Ω is a Greek letter; it is not a set. {Ω} is the set whose only member is the letter Ω. {{Ω}} is the set whose only member is the set whose only member is the letter Ω. If the constraint Ω = {Ω} is imposed on the two-node one-edge graph forcing the two nodes in the regular set to become identical, the result is a hyperset. The constraint need not be imposed only on nodes at adjacent levels. The constraint Ω = {{Ω}} on the three-node two-edge graph forces the top and bottom nodes in the regular set to become identical, turning the regular set into a hyperset.

The graphical representation provides a convenient way to distinguish a set from a hyperset. If the graph includes a set of edges that form a closed cycle, anywhere in the graph, then the graph represents a hyperset. If the graph includes no set of edges that form a closed cycle, anywhere in the graph, then the graph represents a regular set.

Although hypersets are non-contradictory, they are bizarre. The two rightmost graphs in Figure 1 are different graphs that represent the same object. To see that, consider that two sets are equal if they contain identical elements. Start with the fact that Ω = {Ω}. If we nest these two identical objects each in its own container, it remains true that {Ω} = {{Ω}}, since the new containers contain identical objects. However, this does not invalidate the original identity, Ω = {Ω}. Also, since equality is transitive, it must be the case that Ω = {{Ω}}. Since the same nesting trick can be repeated to any depth, it is obvious that we can generate an infinitude of different true propositions about , including Ω = {Ω}, Ω = {{Ω}}, Ω = {{{Ω}}}, and so on. Clearly, the hyperset structure entails much more flexibility than a formalism, with its finitely many true propositions.

There is a temptation to dismiss hypersets because of their flexibility. If they have limitlessly many properties, how could we ever understand them well enough to use as models? However, while hypersets have freedom, they do not have license. Since Ω is provably unique, it must not be the case that Ω = {1}. Neither is it the case that Ω = {2}, nor does any other integer inside the set brackets satisfy the equation. Although many more classes of propositions about are provable, what has been proven so far makes the point. Ω has an infinitude of properties, but is forbidden from another infinitude of properties. As with L'HOPITAL's rule in calculus, the clash of infinities in hypersets leads to finite properties. Hyperset structures are more flexible than regular set structures, but that flexibility is constrained.

In this constrained flexibility, where some things are required and some are forbidden, we have a process that is both understandable and useful. A logical proposition, such as $p \rightarrow q$ is an inferential entailment. A hyperset structure does not merely possess a scattered infinitude of unrelated inferential entailments. The coherence of the hyperset structure imposes an organization on the infinitude of inferential entailments. All the infinitely many propositions of the form $p \rightarrow q$ are organized into an inferential entailment structure. Given a specific hyperset structure, we can determine the inferential entailment structure, and we can ask questions about that structure without the need to consider each of the individual entailments.

Why does this matter? The inferential entailment structure of a hyperset forms a closed loop that winds its way through multiple levels in a hierarchy. The utility of these properties will become apparent later. A hyperset is permitted to have both internal and external semantic referents. It is an inherently ambiguous object, and can be used as a mathematical model of semantic ambiguity.

How does it do that? We recall that an unambiguous formalism is a single finitely-large cluster of propositions about a single external referent. In a hyperset structure the infinitude of propositions can form hierarchically organized systems of several distinct clusters, each containing its own infinitude of propositions, and each corresponding to a different meaning. The structure is similar to KORZYBSKI's semantical *multiordinal terms*, whose main characteristic is that on different levels of orders of abstractions they may have different meanings

The hyperset-based structure serves as a model in the sense that by asking questions about a cluster, we can gain understanding about its referent. BARWISE and ETCHEMENDY demonstrate this in their solution to the Liar Paradox. The problem starts by claiming "This sentence is false." The apparent problem is that if it is true that the sentence is false, does not that very truth negate the falsehood that the sentence asserts? The first point that BARWISE and ETCHEMENDY make is that a semantic sentence is not a logical proposition. Rather, a structure of logical propositions can serve as a model of the sentence. Thus, the initial confusion at interpreting the Liar Sentence stems, as so much confusion does, from ignoring KORZYBSKI's warning that "the map is *not* the territory" (KORZYBSKI 1994, p750).

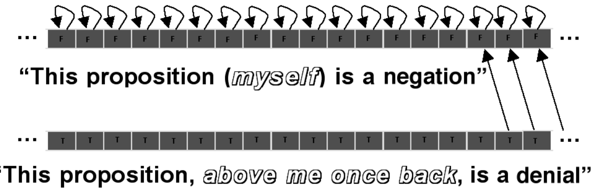

Figure 2. Clusters of propositions in solution to Liar Paradox.

Their hyperset analysis of the Liar Sentence produces two clusters of propositions as seen in Figure 2 (BARWISE/ETCHEMENDY 1987, pp129–138). One cluster is an infinite sequence of propositions each saying, "this proposition (referring to itself) is a negation;" all the propositions in the cluster are false. The other cluster is an infinite sequence of propositions each saying, "this proposition (referring to the proposition in the position one slot back from the corresponding proposition in the other sequence) is a denial;" all the propositions in the cluster are true.

By using a hyperset model to resolve the Liar Sentence into two clusters of propositions, BARWISE and ETCHEMENDY were able to uncover two ambiguities that are the cause of the confusion in the Liar Paradox. The first is the ambiguity in the meaning of *this*; in the top cluster the referent of *this* is internal, in the other cluster the referent of *this* is external. The meanings are totally different. The second ambiguity is rather more dramatic, in the meaning of *false* (BARWISE/ETCHEMENDY 1987, pp164–170). In the top cluster the falsehood referred to is a negation, a proposition that if it were true, it would lead to a contradiction, and so must be false. In the bottom cluster the falsehood referred to is a denial, a proposition that if it were true, it would contradict another true proposition. Thus, the logical propositions that model the Liar Sentence are coherent and straightforward; the paradox vanishes.

Let us not gloss too quickly past an important detail in the preceding paragraph. *The concept of falsehood is ambiguous.* This is *not* a claim that logic should have three truth values, true, negated and denied. For all but a few pathological exceptions, false propositions are simultaneously both negations and denials. However, for those few pathological exceptions, the distinction in meaning between negation and denial must be recognized, or confusion over the nature of the falsehood will ensue.

Recall that formalism is a tool for performing rational thinking by banning ambiguity. If we consider rationality to be non-contradiction, and logic to be the standard by which we assess the absence or presence of contradiction, and if one of the two truth values of logic is itself ambiguous, then we arrive at a surprising conclusion. Intuitively, it is tempting to suppose that if we admit ambiguity into a discussion, then our power to perform rational identification is weakened. However, as we've just seen, the opposite is the case. If ambiguity is not admitted, then the admissible properties of falsehood must be arbitrarily limited, and our power to perform rational identification is weakened thereby. The disallowing of ambiguity that computation imposes is an unreasonable limitation. *When seeking rational answers to difficult questions, the consideration of ambiguity is required rather than forbidden. Furthermore, impredicative mathematics provides a rational model of ambiguity.*

How *Do* We Deal with Ambiguity?

If we construct a small and incoherent model of a process, ambiguities in the model are manifested as contradictions. Conventional wisdom often interprets contradictory alternatives as different candidates for truth. The conventional solution to the problem then degenerates to selecting the least inconvenient alternative and declaring it to be true.

However, this approach ignores a larger truth that might be revealed by a larger model. The way to resolve the seeming contradictions in a small inferential entailment structure is to find a bigger, coherent, inferential entailment structure that accommodates all the cues. An example from art illustrates the point. What are we looking at in Figure 3? We're looking at my picture of the Government of Sweden's picture of M. C. ESCHER's picture of what? Our intuition is to interpret it as a 2-D image of a 3-D static object, since that is what most pictures are.

Such an interpretation is not rational in this case. If it is a representation of a 3-D static object, one has simply to ask, which corner is closest to the viewer. Given the identical "closeness cues" implied by the shading and suggestions of occultation at each corner, each of the three corners looks closer than the other two. Clearly, this is a self-contradictory interpretation of the picture. Whatever is being depicted, the small model that presumes it to be a 2-D representation of a static 3-D object breaks down by contradiction. If the image represents anything at all (perhaps it does not; perhaps ESCHER was merely playing an artistic practical joke), it must be something else.

Figure 3. What is it?

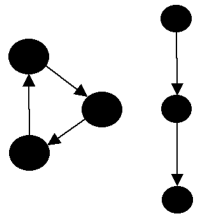

Figure 4. Set and hyperset hierarchies.

Notice that the picture is remarkably similar to the 3-node graphical representation of the prototypical hyperset by the left-hand graph in Figure 4. To appreciate how this works, first consider the three node set represented by the right-hand graph in Figure 4 as a traditional hierarchy. Using the position and orientation of the arrow as hierarchy cues (i.e., the node being touched by the point of the arrowhead is lower in the hierarchy than the node being touched by the tail of the arrow.), we see the middle node to be above the bottom node, but below the top node in the hierarchy.

Using the properties of the arrow as hierarchy cues, we can apply this interpretation to the left-hand graph in Figure 4. First, imagine that the upward directed edge is not present; if it is absent, there is no problem in seeing the remainder of the graph as representing a linear hierarchy as described above. However, as a consequence of ACZEL's proof of the coherent existence of the hyperset, there is nothing preventing including the upward directed edge in the graph.

Given our interpretation of the properties of the arrow as hierarchy cues, it can be seen that any node in the closed-loop graph can be taken as being one position higher than one neighbor and one position lower than the other neighbor in a bizarre hierarchy. This hierarchy of three members, in which it is the case that, for each of the three members, that particular member is exactly in the middle of the hierarchy, is bizarre (counter-intuitive) but non-absurd (logically non-contradictory). As a theoretical matter it is not absurd; the non-contradictory existence of the prototypical hyperset with exactly these properties has been proved by ACZEL. As a practical matter, while such a hierarchy is bizarre, it is no more so than the physical Universe, which is observed to be expanding in all directions away from every observer, no matter where the observer is located.

We also know from the previous section that the hyperset has properties similar to the semantic structure of KORZYBSKI's *multiordinal terms*. The nodes in a hyperset graph can correspond to different meanings of an ambiguous sign where the entire graph corresponds to the entire sign. If we have a sign with three different meanings, is it not reasonable to suppose that each of the three meanings might fit just as well anywhere in a hierarchy of meanings? KORZYBSKI's concept of *multi-ordinal terms*, does not mention this property, but neither does he preclude it. In that case, the three-node hyperset graph in Figure 4 could represent the relationship between the meanings. As with the "expanding from everywhere" property of the Universe, the idea is bizarre, but not absurd.

Again, assuming that ESCHER is not merely having a moment of fun with us learned Doctors of Philosophy, perhaps the comparison of Figure 3 with Figure 4 suggests a larger model in which the picture in Figure 3 is what you get when you try to get an image of semantics by projecting a semantic process onto a syntactic thing. This hypothetical larger model is more satisfying than the smaller incoherent model in an important sense. In this larger model (where the picture is a syntactic representation of semantic process), the shadings and overlaps that we had interpreted as "closeness cues" in the smaller model (where the picture is 2-D image of static 3-D object) now turn out to be "hierarchy cues" between nodes in a semantic process, where it is indeed the case that each node is exactly in the middle of a closed-loop hierarchy formed by three nodes.

Thus, it is demonstrated that a rational solution to the seeming contradiction in cues implied by the small model of Figure 3 is to hypothesize a larger model, in which the contradictions vanish. To hypothesize a claim that Figure 3 is a glimpse at semantics is not final proof that it is. However, the claim is non-contradictory; it is not patently false in the sense of being a negation. It can be treated as tentatively true, but subject to POPPERESQUE falsification. In other words, we must admit the possibility that the claim can be shown to be a denial by the production of an as yet undiscovered valid counter-example. However, the hypothesis is stronger than the claim that Figure 3 depicts a 3-D static object; the weaker hypotheses hypothesis is a negation, self-contradictory or absurd on its own face.

This interpretation of ESCHER's image illustrates how we deal with ambiguity. We construct a model sufficiently large that, within its constraints, the inherent ambiguities do not lead to contradictions. We form hypotheses that resolve seeming contradictions. Computers do not do this; brains do. Hypothesis formation is based on taking ambiguity into account, and by design, computers not only ignore ambiguity, but are inherently incapable of dealing with it. In contrast, as Walter FREEMAN observes, "brains are hypothesis driven" (FREEMAN 2000).

What FREEMAN concludes from decades of observations in salamander brains, FODOR concludes from the obvious context dependency of cognitive processes. According to FODOR, brains perform *abduction*, and computers do not (FODOR 2000, p28). The remainder of FODOR's book addresses the topic of why the various claims that "abduction is computable" are not valid. Those claims all center about hidden ambiguities not noticed by the claimants, and are invalidated when the ambiguities are revealed. Note that the observation that brains perform abduction is not a claim that only brains perform abduction. Abductive behaviors are common in biology. The point here is that abduction is an observable distinction between brains and computers.

The idea of abduction is not new; PEIRCE articulated it over a century ago. "Abduction is the process of forming an explanatory hypothesis; it is the only logical operation which introduces a new idea; for induction does nothing but determine a value, and deduction merely evolves the necessary consequences of a pure hypothesis" (Quoted in DELEDALLE 1990, p60). What is new is the insight discovered independently by the neurophysiologist FREEMAN and the philosopher FODOR, each operating within their nearly unrelated disciplines. *Resolution of ambiguity by abduction is a unique and distinguishing observable signature of incomputable behavior.*

Causation in Mechanisms and Organisms

Having identified a fundamental difference between brain and machine behavior, that brains abduct and machines do not, we find that this answer, as most good answers do, raises more problems than it solves. The effects of abduction are quite dramatic; it resolves seeming contradictions, it does so with a speed and reliability that far outstrip blind guessing, and it creates genuine novelty. What could cause such an amazing effect? Before we can intelligently inquire what causes abduction, we need to understand a bit about how causation operates in organisms.

An effect or event occurs, and we ask, "Why did this happen?" Remarkably, ARISTOTLE's description of cause has withstood 2300 years of criticism and has survived intact. He says, "… causes are spoken of in four senses. In one of these we mean the substance, i.e., the essence (for the "why" is reducible finally to the definition, and the ultimate "why" is a cause and principle); in another the matter or substratum, in a third the source of the change, and in a fourth the cause opposed to this, the purpose and the good (for this is the end of all generation and change)" (ARISTOTLE, Book I, Part 3). The substratum is the *material cause*, that which is changed into an effect. The essence, or the principle which determines the form of the effect, is the *formal cause*. The source of the change is the *efficient (or moving) cause* (distinct from, and not to be confused with the modern concept of efficiency). The purpose and good of the change is the *final cause*.

To the question, "Why a house?" ARISTOTLE answers. "The same thing may have all the kinds of causes, e.g., the moving cause of a house is the art or the builder, the final cause is the function it fulfils, the matter is earth and stones, and the form is the definition" (ARISTOTLE, Book III, Part 2). Note in particular that the moving or efficient cause is identified with the art or the builder, in modern parlance, the "builder's knowhow." The formal cause is the definition; in modern construction this is defined on a blueprint. The material cause is the building stones that were transformed into a house. The final cause is function; it is a place to live.

There is a temptation among modern students of ARISTOTLE to confuse efficient and material cause. Often the flight of a ball is characterized as having been efficiently caused by a kick. In classical mechanics, the flight of a ball is characterized by an ordered set of transformations through a succession of states, each fully described by a state vector consisting of displacement and velocity of the ball at a given instant. The effect is the current state; the process of flight is the transformation to the current state from the prior state. Thus the prior state is the material cause, that which was transformed. The first state (or initial conditions that were established by the kick) in the set of transformations, is clearly the first *material cause* in the progression of states.

Why does the ball, given a particular set of initial conditions, follow a particular trajectory? It is forced or constrained to do so by external forces such as gravity and aerodynamic drag. It is these constraints which are the "source of change" in the trajectory. In other words, these constraints, often summarized in a differential equation and characterized as a "dynamical law" are the *efficient cause* in the progression of states.

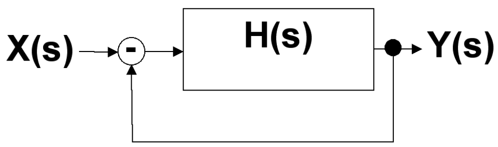

Figure 5. A linear servomechanism.

Causation in a mechanism is illustrated by the linear servomechanism in Figure 5. The *effect* or output, or next state, of the servo is characterized by *Y(S)*. What causes *Y(S)*? One of the causes is *X(S)*, a set of present and previous states also known as initial conditions; what the servo does is to transform states *X(S)* into state *Y(S)*. *X(S)* is a material cause. In fact, engineers would call the servo a "causal system" if and only if *X(S)* does not contain "information from the future." In other words, the engineering notion of causality is limited to material cause. Thus, despite the fact that the servo features a closed loop of causal entailment, all the entailments in the loop are material causes. Note that this is in keeping with the thermodynamic closure of mechanisms; the system neither changes nor is changed by its environment. It is closed to material cause.

Obviously, the transformation of *X(S)* into *Y(S)* does not proceed willy-nilly. The process is constrained by $Y(S) = H(S)X(S)$, where *H(S)* is technically termed the *transfer function*. The transfer function is a Law of Behavior, or a description of how the servo constrains the transformation process. The transfer function serves the role of efficient cause.

The transfer function is determined by the interaction of the physical topology of the servo and the constraints inherent in reality that are often characterized as the Laws of Nature (MAXWELL's equations, NEWTON's Laws, and so on). Note that if we ask where the efficient cause came from, we find that it is externally entailed. An engineer, knowing the laws of nature, synthesizes a topology, which constrained by the laws of nature must lead to an efficient cause that is described by *H(S)*. The important point is that the efficient cause of the mechanism is externally entailed.

A mechanism need not be designed by the Hand of Man in order to have an efficient cause. In the case of a naturally occurring mechanism, such as a planetary orbit, a particular topology is given (why or how it is given is irrelevant), and that topology, being governed by the Laws of Nature produces an efficient cause. In either case, whether entailed by the Hand of Man, or by some invisible hand in nature, the crucial point is that in mechanisms, efficient cause is externally entailed. Mechanisms are open to efficient cause.

To account for the specific form of a specific output, there must be a constraint on efficient cause. That constraint is *formal cause*. This is an abstraction of the output, rather than the output itself. "The form is the definition" (ARISTOTLE, Book III, Part 2). In a servo, the form or definition is seen in the parameters that would show up as specific values in the transfer function and would affect specific constraints on the output of the particular system (ROSEN 1987).

By far the most controversial of the ARISTOTELIAN causes is final cause, the answer to the question, "What purpose is the effect for?" For man-made productions, this may not be a big problem. A servo achieves whatever purpose its designer chooses; beyond that, it has no effect on final cause and is unaffected by it. In the case of naturally occurring mechanisms, such as planetary orbits, questions of final cause are disallowed. The perfectly mechanistic Universe described by LAPLACE is not *for* anything; the Universe is simply a historic accident, unfolding without goal or purpose. The crucial point is that no mechanism, natural or man-made, produces its own final cause.

"That a final cause may exist among unchangeable entities is shown by the distinction of its meanings. For the final cause is (a) some being for whose good an action is done, and (b) something at which the action aims; and of these the latter exists among unchangeable entities though the former does not" (ARISTOTLE, Book XII, Part 7). ARISTOTLE defines final cause as goal seeking, and for a changeable entity, that goal is the good of some being. Can the changeable entity be an organism? Can the being whose good the entity seeks be its very self? ARISTOTLE did not think so. "And so, in so far as a thing is an organic unity, it cannot be acted on by itself; for it is one and not two different things" (ARISTOTLE, Book IX, Part 1).

However, modern observations show that organisms do exhibit self-determined goal-seeking behavior. "I will begin by giving a name (intentionality) to

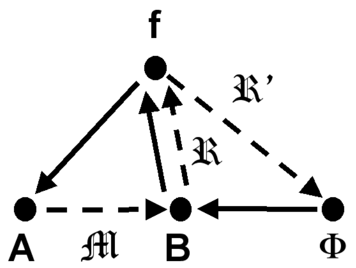

Figure 6. Causal entailment structure of an organism.

the process by which goal-directed actions are generated in the brains of humans and other animals" (FREEMAN 1999a, p8). FREEMAN goes further than merely admitting goal-seeking behavior to the discussion; he sees it as crucial. "Intelligent behavior is characterized by flexible and creative pursuit of endogenously defined goals" (FREEMAN 1999b, p185). FREEMAN is not the only physician to observe this property in living processes. "A simple organism made up of a single cell, say, an amoeba, is not just alive but bent on staying alive … the form of the intention is there, nonetheless, expressed by the manner in which the little creature manages to keep the chemical profile of its internal milieu in balance, while around it, all hell may be breaking loose" (DAMASIO 1994, p136).

In processes of life and mind, self-determined goal-seeking behavior is consistently observed by reliable witnesses. Yet, ARISTOTLE did not believe that an entity could simultaneously be both cause and effect. From this perspective, shared by most modern scientists, he missed the fundamental difference between mechanisms and organisms.

To appreciate how deep that difference is, recall the structure of causation for the servomechanism shown in Figure 5. A *causal entailment* ($a \rightarrow b$) is a specific instance of "event *a* causes event *b*." The *causal entailment structure* of a servomechanism is nothing more than a single-level closed-loop of entailments of material causes.

Contrast the causal entailment structure of a servomechanism with the one of the simplest conceivable causal entailment structures of an organism, as shown in Figure 6. In the diagram in Figure 6, originally devised by Robert ROSEN, the nodes represent events and the directed edges represent causal entailments. The dashed edges are material causes, and the solid edges are efficient causes. The organism is a physical *process* consisting of several different causally entailed subprocesses (ROSEN 1991, pp248–253). (Note: The causal entailment structure is fully consistent with MATURANA and VARELA's concept of *autopoiesis* (MATURANA/VARELA 1981). ROSEN has filled in some blanks as to how the causal entailments interact.)

Subprocess $\mathfrak{M}$ is metabolism. Living processes differ from the thermodynamic closure of mechanisms. "Living organisms continually obtain energy and materials from the external environment and eliminate the end products of metabolism. Being open systems, they are not subject to the limitations of the second law of thermodynamics" (MAYR 1997, p22). From Figure 6, we see that in subprocess $\mathfrak{M}$, the material cause, A, input from the outside world, is transformed into the effect B, the substance and organization of the organism, and the transformation is regulated by efficient cause f. The important point to note is that the organism is open to material cause.

Since the organism is open to material cause, it receives more than nutrients from the environment. It is open to an enormous array of environmental insults that can disrupt its internal subprocesses, including metabolism, $\mathfrak{M}$. To survive these environmental insults, "it continually self-repairs" (MARGULIS 1995, p17). The repair subprocess, $\mathfrak{R}$, keeps efficient cause f in repair. From Figure 6, we see that in subprocess $\mathfrak{R}$, the material cause, B, is transformed into the effect, f, and the transformation is regulated by efficient cause ϕ.

Of course, the repair subprocess, $\mathfrak{R}$, is no less vulnerable to environmental insult than the metabolism subprocess, $\mathfrak{M}$. How do we keep the repair process in repair? ROSEN showed that we could do so by constructing a replication subprocess $\mathfrak{R}'$. $\mathfrak{R}'$ produces ϕ, a replica of f. The material cause, f, is transformed into the effect, *f*, and the transformation is regulated by efficient cause *B*.

We see the unfortunate possibility of an infinite regress developing. $\mathfrak{R}$ repairs $\mathfrak{M}$. $\mathfrak{R}'$ repairs $\mathfrak{R}$. However, the organism is finite, having a bounded identity. "Islands of order in an ocean of chaos, organisms are far superior to human-built machines." (MARGULIS 1995, p17). Such an "island" is incompatible with an infinite regress of repair subprocesses. The inventiveness of ROSEN's causal entailment structure is that it keeps everything in repair without recourse to an infinite regress.

Except for A, the externally imposed material cause, all the nodes in Figure 6 are internally generated effects, and all serve as the efficient cause of some other effect within the organism. Stated slightly differently, all the efficient causes, or constraints on the behavior of the organism, are themselves effects caused by subprocesses within the organism. In other words, an organism is "closed to efficient cause."

This structure of causal entailment has some peculiar properties. *B* is the efficient cause of ϕ. ϕ is the effcicient cause of f. f is the efficient cause of *B*. We see this in Figure 7, and we note a remarkable fact, like the structure of multi-ordinal terms in semantics, this structure is identical to a hyperset hierarchy. In other words, each efficient cause is a member of a hierarchy, and each is exactly in the middle of the hierarchy. We would be tempted to dismiss this structure as incoherent, except that ACZEL has proved that it *is* coherent.

Most significantly, this structure shows why ARISTOTLE's argument against self-causation is needlessly restrictive. He saw causes as operating in a linear hierarchy, *B* causes ϕ causes f, and argues that B could not be both the cause and the thing caused. The fact that there is a non-contradictory way to form a multi-level closed-loop hierarchy, such that *B* causes ϕ causes f causes *B*, is a 20th century discovery, and was unknown to him. Nevertheless, the hyperset hierarchy allows us to relax ARISTOTLE's restriction in a coherent and non-arbitrary way. *B* can be both cause and effect in the same hyperset-like hierarchy, and there is no difficulty in seeing *B* as being self-caused.

The hierarchy of causation appears to move in two directions. Bottom-up causation is where the parts drive a whole structure of parts, and traditional reductionism has no difficulty with this concept. Top-down causation is where the whole structure drives the action of the parts, and this is dismissed as mysticism. However, in a hyperset-like closed loop of causation we note that while the direction of the causation around the loop is indeed unidirectional, we can move from any node to anywhere else we like. f causes *B*, bottom-up. However, it is also the case that B causes ϕ causes f, also bottom-up, but it creates exactly the same effect as if *B* causes f, top-down.

The hyperset-like multi-level closed-loop hierarchy of efficient cause gives us a way of obtaining seemingly impossible effects through a progression of bottom-up moves the long way around the loop, provided the loop has finitely many nodes. As is evident from Figure 7, we can transit the loop bottom-up from any given position until we end up exactly one position down from where we started; this gives us top-down behavior, final cause, or the whole operating on the parts. We can also make the one last move that causes us to end where we started, thus

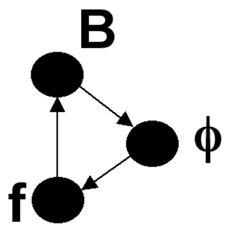

Figure 7. Structure of efficient cause in organisms.

having any cause in the loop serving as its own cause.

Juxtaposing both of these facts, an entity can serve as its own final cause. To paraphrase ARISTOTLE, the entity acts for the good of the agent of its final cause. The bizarre result of the hyperset-like structure, that the entity can be the agent of its own final cause. Thus, it follows that final cause can operate for the organism's own good, just as FREEMAN and DAMASIO have observed. The reason that ARISTOTLE excluded this possibility was that he was unaware of hyperset-like hierarchical structure of causal entailment.

Operating from mathematical principles, ROSEN determined that an organism must be semantic, constructing knowledge for its own good, internally self-organizing (making up all its own efficient causes), context dependent (open to material cause), and performing all this in a multi-level closed-loop entailment structure. DAMASIO determined substantially the same thing through his medical observation of brains. "The entire construction of knowledge, from simple to complex, from non-verbal imagetic to verbal literary, depends on the ability to map what happens over time, *inside* our organism, *around* our organism, *to* and *with* our organism, one thing followed by another thing, causing another thing endlessly" (DAMASIO 1994, p185).

This discussion has accumulated quite a few differences in causation between an organism and a mechanism, and it is useful to pause a moment and compare them. A mechanism is closed to material cause, open to efficient cause, and operates for the good of some external agent of final cause (supposing that it has one). A mechanism may have a closed loop of causal entailment, but the entailments in the loop are a progression of material causes. The kinds of causes in a mechanism are all separable.

In contrast, an organism is open to material cause, closed to efficient cause, and its internally created final cause drives it to operate for its own good. It must have a closed loop of causal entailment, but the entailments in the loop form a bizarre hierarchy of efficient causes, where each cause appears to be in the middle of the hierarchy. At least some of the causal entailments are inseparably entangled, simultaneously serving as material and efficient causes of different effects.

Recall that FREEMAN says that this is exactly what brains do. "Intelligent behavior is characterized by flexible and creative pursuit of endogenously defined goals." Notice particularly that he calls this multi-level self-referential process of causation by the medical term, *endogenous*. The term *endogenous system* is conventionally used in medicine to describe the property of a system that grows from within itself, or makes itself up as it goes along (CLAYMAN 1989). The term is used for a similar concept in economics (GAVIN/KYDLAND 1999).

As I have suggested in other comments intended for engineers, I also suggest to the students and practitioners of cognitive neuroscience (KERCEL 2001). *Endogenous system* is the most descriptive and least confusing term for describing a natural system that is distinguished by a closed-loop multi-level structure of efficient cause. I note that the causal entailment structure of an endogenous natural system is similar to the inferential entailment structure in an impredicative mathematical system.

What Causes Abduction?

Thus far, we have observed three similar bizarre hierarchies, the structure of multi-ordinal terms in semantics, the nesting structure in hypersets, and the structure of efficient cause in organisms. In each structure, each functional component is a member of a hierarchy, and each is exactly in the middle of the hierarchy.

Can we use this similarity to construct a rudimentary model of abduction? The hyperset inferential entailment structure has the property of being fully determined in a peculiar way. It has *an infinitude* of propositions, but every inferential entailment of the form "proposition p implies proposition q," is itself implied by other inferential entailments in the inferential entailment structure. The similarity between hypersets and endogenous causation suggests a hypothesis that the same sort of determinism applies in an endogenous natural system, that every causal entailment of the form "event a causes event b," is itself an effect of other causal entailments in the causal entailment structure.

This idea, that there are no uncaused effects, has credence in some schools of philosophy. The similarity between hyperset propositions and endogenous causal entailments demonstrates that the idea is not contradictory on its own face. It might still be possible to show that it is a denial. In other words, it becomes a worthy candidate for POPPERESQUE falsification. It is reasonable to regard it as being hypothetically valid, subject to being proven false by showing that it is contradicted by a valid counterexample.

Does ARISTOTLE provide the counterexample? He speaks of an uncaused first cause. However, he needed an uncaused first cause only because his hierarchy of cause was linear. In a multi-level closed-loop hierarchy, self-caused causes are allowed, and uncaused first causes or unmoved prime movers are not necessary. In the absence of that necessity, ARISTOTLE provides no other evidence of its existence.

Nevertheless, he would still object to the notion of an infinitude of causal entailments. "But if the kinds of causes had been infinite in number, then also knowledge would have been impossible; for we think we know, only when we have ascertained the causes, that but that which is infinite by addition cannot be gone through in a finite time" (ARISTOTLE, Book II, Part 2). His actual objection to infinitudes was that they cannot be counted. However, CANTOR showed that there is no need to count to infinity in order to consider infinitudes; it is rational to discuss finite representations of infinitudes, and to prove non-contradictory propositions about them (ECCLES 1997, pp176–178). Thus, ARISTOTLE's actual objection to infinitudes has been falsified by CANTOR.

There might still be a problem. Does not an infinitude of causal entailments require an infinitude of entailing events in reality? Is this not impossible in a finite Universe? Clearly, an infinitude of real events in neither possible nor necessary. ARISTOTLE saw causal entailments existing as potentials before they took ontological form. "Further, matter exists in a potential state, just because it may come to its form; and when it exists actually, then it is in its form. And the same holds good in all cases, even those in which the end is a movement" (ARISTOTLE, Book IX, Part 8). He also saw that just because an event had a potential for occurring, there was no specific requirement that it must occur. "For the potency is prior to the actual cause, and it is not necessary for everything potential to be actual" (ARISTOTLE, Book III, Part 6). In other words, more causal entailments exist as potentials than are entailed as caused events. Furthermore, there appears to be no rational objection to the possibility (suggested by the similarity to the hyperset inferential entailment structure) that the number of those potential causal entailments is limitless.

There is one other crucial similarity between ARISTOTLE's view of causal entailment and hyperset-based inferential entailment. Just as some propositions are forbidden by the inferential entailment structure, ARISTOTLE saw that some events could and must be precluded by the causal entailment structure. "Nothing that is incapable of being comes to be" (ARISTOTLE, Book III, Part 6). Again, there appears to be no rational objection to the possibility (again, suggested by the similarity to the hyperset inferential entailment structure) that the number of events forbidden by the causal entailment structure of reality is limitless.

The hypothesis is that events in reality are fully caused in the same sense that propositions in a hyperset structure are fully determined. The point of this discussion is that the hypothesis has not been falsified by the objections that have been raised thus far. Following POPPER's principle of falsification (which he applies to all hypotheses about reality) we can treat the hypothesis as true, subject to the risk that it might be falsified in the future.

We've already seen that the determinism that is the subject of this hypothesis is of a very peculiar character. Recall that in the hyperset model, the inferential entailment structure both permits the hyperset to do an infinitude of things, and forbids it from doing another infinitude of things. The clash of these two infinitudes leads to a structure of finitely many clusters of infinitudes of propositions. When hypersets are used to model semantic signs, the clusters of propositions form a hyperset structure whose nodes in correspond to different meanings of the sign.

It becomes not much of a stretch to suppose that a similar thing happens in the dynamics of an endogenous process. Perhaps the structure agglomerates to finitely many infinite clusters of causal entailments, each of which serves as the cause of a potential occurrence, just as ARISTOTLE described potential. Thus, we find ourselves with several entailed possibilities, of which only one actually occurs. Nevertheless, whichever occurred, it was fully caused. (It is beyond the scope of this model to explain why one possibility occurs instead of another. Likewise, it is beyond the scope of the physicists' model of radioactive decay to explain why a

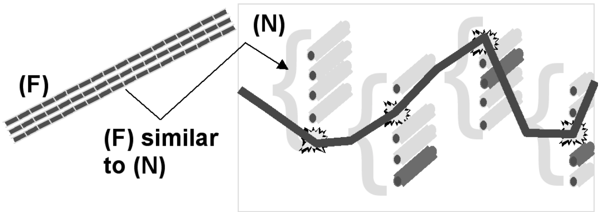

Figure 8. Working through a path of entailed possibilities.

particular atom decays next instead of another.) In the case of mental processes, the event is manifested as a choice, fully caused, but not fully predictable.

Now, the causation becomes even a bit more bizarre. In endogenous causation, the causal entailment structure is itself caused, and the occurrence of a caused event also causes an update of the entire causal entailment structure. The subsequent event is caused by a different (but possibly not much different) agglomeration of finitely many infinite clusters of causal entailment. This progression of a process through reforming clusters of entailed potential events is shown in Figure 8. *F* is the structure of infinite clusters of propositions representing nodes in the hyperset model. *N* is the corresponding set of infinite clusters of causal entailment representing potential events. For each structure of potentials, only one of the events occurs in reality, and the model provides no information of which one. As the meanings of an ambiguous sign are all dead center in the semantic hierarchy, all the potential events are dead center in the hierarchy of potentials.

This concept lends itself to a remarkable degenerate case. Suppose that for a given causal entailment structure the number of "clusters of possibility" is exactly one. Also suppose that it is a property of the entailment structure that when the entailed event occurs, the reformed structure is identical to its predecessor. In this degenerate case, we have recovered the absolutely rigid causal entailment structure of classical and relativistic mechanics.

Recall that in discussing processes of life and mind, SCHRÖDINGER allowed the possible necessity of a super-physical law. The causal entailment structure depicted in Figure 8 indicates the causal entailment structure of just such a biologically-based super-physics. Would this serve as the causal basis for super-physical "Laws of Biology" that include the laws of classical and relativistic physics as a degenerate case? It is quite possible.

(The curious reader may wonder why quantum mechanics is excluded from this discussion. From the standpoint of causation, there are no Laws of Quantum Mechanics. At its foundation, quantum mechanics is strictly epistemological. It denies causation, and ignores the paradoxes that flow from that denial (JAYNES 1989)).

There is one other consequence of the causal entailment structure shown in Figure 8 that is breathtaking in its implications. Causal entailments form potentials, and not the events themselves. Most of the potentials never actually cause events. Thus, while only finitely many events are actually caused, involving only a finite amount of matter and energy, the amount of causal entailment available to the structure is limitless just as the amount of semantic inferential entailment in the hyperset model is limitless. If there is limitless causal entailment available from which the entailment structure may reform after an event occurs, then there is no necessity that the new structure be identical to any causal entailment structure that has ever occurred in the past.

In other words, the possible occurrence of *novelty* is inherent in the entailment of the entailment structure and any novel event that does occur is fully caused. Thus, based on the similarity with semantic novelty flowing from the inferential entailment structure of a hyperset model, it appears reasonable that we might expect that novelty can flow directly from the causal entailment structure of an endogenous natural system. Novelty is *caused* by the endogenous causal structure of reality; it is not the result of a magical *élan vital*, an equally magical uncaused quantum effect, or an even more mysterious "emergence."

If novelty is caused, is it not determined? Yes, determined means caused. If it is determined, is it not known in advance, thus making it not novel at all? No. Only in reductionism, where the model is a complete description of the process, is determinism identical with knowing in advance.

In contrast, an impredicative model is *similar* to an endogenous system, but the two entailment structures never fully correspond. Given a high degree of correspondence between the model and the system, many (but not all) behaviors may be known in advance but an unanticipated event can still occur. For example, the mental model of a border collie in the mind of a shepherd is very similar to the actual behavior of the dog. However, that does not preclude the dog from doing something new, and often better than what the shepherd wanted him to do.

After what seems a long digression, we can now describe the cause of abduction. When seemingly contradictory sensory cues are fed into the brain from the outside world, a dangerous situation is perceived, and it drives a feeling of distress. Since, as DAMASIO says, the organism is intent on staying alive, or as FREEMAN says, its intentionality drives it to seek its own good, the endogenous (closed-loop multi-level) causal entailment structure entails a final cause, intentionally operating in the direction of the organism's own self-determined good.

The endogenous causal entailment structure encompasses an infinitude of causal entailments. The structure of those entailments changes with every event, and enables a choice of possibilities for the subsequent event. There is nothing in the structure to preclude it from adding something new to the entailment structure and to the list of possibilities that it entails. Hence, the entailment structure can and does become bigger as it operates.

In the case of conflicting sensory cues, the events entailed by the structure are semantical abstractions of the sensory data. Eventually (often very soon) the caused structure of semantics grows big enough to become similar to the external event sending the seemingly contradictory cues to the mind. This can result in a novel hypothesis that will resolve the seeming contradiction. The abducted hypothesis is then inductively tested by a process that FREEMAN calls preafference and reafference until one is found that, when tested, actually does resolve the conflict. The resolution produces what DAMASIO calls the feeling of knowing. *The organism's causal entailment structure produces everything that is needed to quickly abduct a novel hypothesis to resolve the seeming contradiction.*

Conclusion

Ironically, few of these concepts are particularly new. What is new in this paper is the observation that semantic ambiguity, hyperset nesting, and the organization of efficient cause in organisms have exactly the same structure, a multi-level closed-loop hierarchy in which every one of finitely many nodes is exactly in the middle. In hyperset structures, each node corresponds to an infinitely large cluster of inferential entailments. In organisms, each node corresponds to an infinitely large cluster of causal entailments.

By the simple artifice of taking ambiguity into account in our reasoning, and admitting into the discussion a mathematics large enough to describe semantic ambiguity, we can understand many of the causal entailments of the endogenous physical processes of life and mind. In short, the principles of endogeny, impredicativity, and the non-identical similarity between them lay the foundations for the discovery of causally-based laws of biology and psychology so general in their scope that classical and relativistic physics fall out as a degenerate case of them.

Could the occurrence of this degenerate case be a clue that the causal entailment structure hypothesized in this paper might serve as a basis for SCHRÖDINGER's super-physical law? Could the corresponding inferential entailment structure in the world of hypersets serve as the "some other tool" that BROOKS says "we need to understand in order to really describe what's going on?" If we are to gain more than a superficial understanding of processes of life and mind, we really must do as BROOKS suggests, and move beyond computation. *Abduction provides the specific impetus to make this conceptual move; it is a readily observable effect that never occurs in computers, but commonly occurs in brains.*

Acknowledgments

I thank Dr. John COLLIER of the Konrad Lorenz Institute for Evolution and Cognition Research and Dr. Edwina TABORSKY of Bishop's University for their helpful review and comments. I also thank the Government of Sweden for putting a specimen of M. C. ESCHER's art on a postage stamp.

Author's address

Stephen William Kercel, New England Institute, 2 Brian Drive, Brunswick Maine 04011, USA.
Email: kercel1@suscom-maine.net

References

Aczel, P. (1988) Non-well-founded sets. CSLI Lecture Notes, No. 14: Stanford CA.

Akansu, A. N./Haddad, R. A. (1992) Multiresolution signal decomposition. Academic Press: New York.

Aristotle (350 BC) Metaphysics. (Translated by W. D. Ross). Retrieved from the World Wide Web at http://classics.mit.edu/Aristotle/metaphysics.14.xiv.html

Bach-y-Rita, P. (1995) Nonsynaptic diffusion neurotransmission and late brain reorganization. Demos Publications: New York.

Barwise, J./Etchemendy, J. (1987) The liar. Oxford University Press: Oxford.

Brooks, R. (2002) Beyond computation: A talk with Rodney Brooks. Edge, June 5. http://www.edge.org/3rd_culture/brooks_beyond/beyond_ index.html.

Chomsky, N. (2000) Language and thought (Fifth Printing). Moyer Bell: Wakefield RI. Originally published in 1993.

Clayman, C. B. (1989) The American medical association encyclopedia of medicine. Random House: New York.

Damasio, A. R. (1994) Descartes' error. Avon: New York.

Deledalle, G. (1990) Charles S. Peirce: An intellectual biography. John Benjamins Publishing Company: Amsterdam.

Eccles, P. J. (1997) An introduction to mathematical reasoning. Cambridge University Press: Cambridge.

Fodor, J. (2000) The mind doesn't work that way. MIT Press: Cambridge MA.

Freeman, W. J. (1999a) How brains make up their minds. Weidenfeld and Nicholson: London.

Freeman, W. J. (1999b) The neurodynamics of intentionality is the basis of intelligent behavior. In: Dagli, C. H. et al. (eds) Intelligent engineering systems through artificial neural networks, Vol. 9. ASME Press: New York, pp. 185–191.

Freeman, W. J. (2000) Comment at the Conference banquet, November 7. Artificial Neural Networks in Engineering 2000, St. Louis.

Gavin, W. T./Kydland, F. E. (1999) Endogenous money supply and the business cycle. Review of Economic Dynamics 2(2):347–369.

Goldrei, D. (1996) Classic set theory. Chapman and Hall: London.

Jaynes, E. T. (1989) Clearing up the mysteries—the original goal. In: Skilling, J. (ed) Maximum Entropy and Bayesian Methods. Kluwer Academic Publishers: Dordrecht, pp. 1–27.

Kercel, S. W. (2001) Does incomputable mean not engineerable? In: Dagli, C. H. et al. (eds) Intelligent engineering systems through artificial neural networks, Vol. 11. ASME Press: New York, pp. 461–467.

Kleene, S. (1950) Introduction to metamathematics. Van Nostrand: Amsterdam.

Knuth, D. E (1973) The art of computer programming Vol. 1 (2nd ed.). Addison Wesley: Reading MA.

Korzybski, A. (2000) Science and sanity (5th ed. 1994, 2nd printing). Institute of General Semantics: New York.

Margulis, L./Sagan, D. (1995) What is life? University of California Press: Berkeley CA.

Maturana, H. R./Varela, F. J. (1981) Autopoiesis and cognition: The realization of the living. Boston Studies in the Philosophy of Science, Vol. 42. D. Reidel Publishing: Boston.

Mayr, E. (1997) This is biology. Harvard University Press: Cambridge MA.

Picard, J. R. W. M. (1993) Impredicativity and turn of the century foundations of mathematics: Presupposition in Poincaré and Russell. Doctoral Dissertation in Philosophy. MIT: Cambridge MA.

Rosen, R. (1987) Some epistemological issues in physics and biology. In: Hiley, B. J./Peat, F. D. (eds) Quantum Implications: Essays in Honor of David Bohm. Routledge, Kegan and Paul: London, pp. 327–341.

Rosen, R. (1991) Life itself: A comprehensive inquiry into the nature, origin and fabrication of life. Columbia University Press: New York.

Schrödinger, E. (2000) What is life? Cambridge University Press: Cambridge. Originally published in 1944.

Sipser, M. (1997) Introduction to the theory of computation. PWS Publishing: Boston.

Von Neumann, J. (1966) Theory of self-reproducing automata, Fifth Lecture. Edited and completed by A. W. Burks. University of Illinois Press: Urbana.

Von Neumann, J. (2000) The computer and the brain. Yale University Press: New Haven CT. Originally published in 1958.

Leib Litman/Arthur S. Reber

On the Temporal Course of Consolidation of Implicit Knowledge

Abstract

The authors examine a variety of still unresolved questions concerning the consolidation of implicit or unconscious knowledge. First, does knowledge acquired independently of intentions to learn and awareness of what was actually learned need to consolidate? Second, if so, is the temporal course of consolidation similar to that obtained from the more usually studied conscious memories? Third, do the various kinds of implicitly acquired knowledge show different time patterns? Fourth, is sleep important in fixing new knowledge? Answers are slowly emerging and some pilot data that speak to these issues are presented.

Key words

Implicit learning, memory consolidation, sleep.

Introduction

In this paper we explore a topic that has received virtually no attention in the literature in the cognitive neurosciences, the consolidation of implicit knowledge. The two distinct focuses here, consolidation and implicit (or unconscious) knowledge, independent of each other, have not exactly been strangers to cognitive science. On one hand, consolidation has been the subject of intense scrutiny ever since it was recognized that organisms cannot function effectively unless they have some mechanism whereby information about events can be retained after those events have ended (MÜLLER/PILZECKER 1900; SCHACTER 2001). On the other hand, implicit learning and implicit memory have been among the more intensely researched topics in cognitive psychology over the past several decades. But the integration of the two domains has been virtually ignored.

What appears to have happened is that as researchers focused more and more on the issue of consolidation, the topic was pursued without regard to the nature of the possibly distinct learning processes being recruited. Consequently, any potentially distinct neuroanatomical mechanisms that mediate the laying down of particular memories were ignored. The focus shifted to exploring such problems as the biomolecular details of synaptic change, the time it took for such processes to become complete, the kinds of interfering events that could disrupt the process, and the like. Lost during this period of research was any concern about the possibility that different kinds of learning might undergo distinct forms of consolidation or that different kinds of knowledge might be handled by distinct neural architectures.

Recent work in learning and memory, however, strongly suggests that behaviorally distinct forms of acquisition and representation exist and that they are served by anatomically distinct neural systems. The natural question that almost asks itself at this point is whether or not these separate memorial systems have different neural mechanisms that operate in the consolidation of their respective representations and whether these differences may be manifested by different time courses in consolidation. Since the approach we take here focuses on the forms of knowledge that are acquired largely independent of consciousness, a short overview of the topic of *implicit learning* is in order here.

Implicit Learning

By "implicit learning" we mean the process of acquisition of knowledge that takes place largely independent of the learner's awareness of either the process of learning or the knowledge ultimately attained (see REBER/ALLEN/REBER 1999 for an overview and any of several contributions to STADLER/FRENSCH 1998 for additional details). However, this notion of unconscious or "implicit" functioning has wider use in contemporary cognitive psychology. In particular, it is also found referring to such processes as, (a) perception, as in "subliminal" perception where sensory information is picked up without awareness, (b) memory, where knowledge is held and can affect behavior although the individual is unable to consciously recall or recognize the material, and (c) motor or procedural learning where the details of actions necessary to produce a fluid motion have become automatized and are not known consciously. What characterizes these implicit processes, in a nutshell, is that they are not *explicit*, the operating processes and resulting mental contents are, by and large, not available to conscious introspection and they spin themselves out largely outside of the awareness of the behaving individual.

An example may help here. In the standard (explicit) memory study fairly obvious and easily detectable stimuli (such as a list of common words) are presented to subjects who are asked to commit them to memory and then later to (consciously) recall, or perhaps recognize, them. The procedure is refereed to as "explicit" because people are consciously aware that they are remembering something, and also because they know exactly what it is that they remember and their recall takes place within the spotlight of consciousness.

However, just because people cannot recall something does not mean that they have completely forgotten it, as most of us know from all-to-common daily occurrences. These, so called *implicit memories* show up readily in what are called fragment completion or stem completion tests, (SLOMAN et al. 1988; SQUIRE/KNOWLTON 1995; TULVING 1983). On the fragment-completion task, instead of having subjects recall words such as "ALLIGATOR" from an initial (learning) list they are asked to complete a word fragment such as A_LIG___R. Participants are generally quite good at this task even—and this is the interesting finding—in the absence of an ability to recall the actual word from the initial list. Thus, some memories are consciously accessible, and others influence performance while remaining unavailable for conscious inspection. More about implicit memory later.

Motor, or procedural, learning is also of interest because, like implicit memory, individuals learn to carry out complex motor activities in a smooth and efficient manner with little or no awareness of the actual pattern of actions engaged in (SHADMEHR/HOLCOMB 1997). People typically do not know exactly what actions are involved in balancing a bicycle, in driving a car, typing on a computer, or carrying out of any of a number of other such tasks which are performed rapidly, smoothly and automatically. One of the interesting and, for our purposes, relevant aspects of motor learning is that in order for an individual to express these skills in a fluid manner they must have some kind of internal representation that captures the full scope of the procedure.

As we will argue below, when knowledge becomes well learned and robust, it is, in effect, knowledge that has been *consolidated*. We will be operating from the position that the establishment of all epistemic content, whether the knowledge is sensory, motoric, or cognitive must be accompanied by some form of neurological change that enables it to become (relatively) fixed. A good example of what happens when such knowledge has become part of an individual's motor programs is to observe what happens when an individual goes to pick up a milk bottle that is actually empty but has been painted to look full. The actions are awkward and inappropriate because the established sensorimotor schema is set for a container of a specific weight which is undermined by the empty bottle (SHADMEHR/HOLCOMB 1997).

These two topics, implicit memory and sensorimotor learning, have received the most attention from researchers and what little we know about the consolidation of unconsciously represented knowledge comes from these two areas of study. However, there are reasons for suspecting that generalizations to representations that emerge from standard implicit learning procedures may be unwarranted. First, because of the standard methods in use when implicit memory is examined, the mental representations established are not necessarily based on new knowledge. For example, the subjects in a typical implicit memory experiment already know words like ALLIGATOR when they come in for the experiment. Hence, the kind of consolidation that is operating here carries a large *episodic* element and may not be the same as that in an implicit learning experiment where new knowledge is actually acquired. This point is important and we'll have more to say on it later.

Second, the typical motor learning experiment is relatively free of cognitive processes. The canonical procedure uses sensorimotor tasks such as learning a sequence of movements of a robot-controlled arm in arbitrary sequences or learning how to balance on a highly sensitive beam, acts which have few if any cognitive, symbolic elements. It strikes us as simply unwise to accept generalizations across such disparate domains without further explorations. Implicitly acquired information may show very different patterns of consolidation—it is a classic case of an empirical question.

The type of implicit functioning that is of central concern here is implicit learning. As noted, implicit learning occurs when people come to be able to function in complexly structured stimulus domains without the intention to learn about the underlying structure, and where the nature of the learned material is not available for conscious inspection (REBER 1993). The acquisition of a natural language is an example of this type of implicit learning. Each of us comes to know (implicitly) the underlying rules and patterns of linguistic communication and we use them every time we talk or listen although no one knows these linguistic rules explicitly (PINKER 1994; REBER 1993). A similar tale can be told about such processes as socialization and acculturation. Much of the individual's lack of conscious knowledge about such learned material can be attributed to the underlying statistical complexity of the target environment. And, in fact, even in controlled experimental settings, such as those discussed below, when stimulus displays are highly complex, top-down, consciously modulated processes tend to be counterproductive (REBER 1976; REBER et al. 1980).

In recent years a number of procedures have been developed for studying implicit learning in controlled laboratory settings. The most commonly used are the artificial grammar learning (AGL) task and the sequential reaction time (SRT) task, although a number of others such as the hidden covariation task (LEWICKI 1986) and the production control task (BERRY/BROADBENT 1984) are also found. In the studies we'll report on here, we used the AGL and the SRT tasks since, as will become clear, they lend themselves rather neatly to exploring the consolidation issue.

Artificial grammar learning (AGL). The standard AGL task consists of a learning phase and a testing phase. During learning, participants memorize a series of letter strings that are constructed according to a rather complex set of rules (the artificial grammar). During testing, subjects are asked to classify novel letter strings as to whether they conform with the rules of the grammar or violate them in some fashion.

Two sample AG's are presented in Figures 1 and 2 along with some examples of grammatical letter strings that each generates. Formally, each of these systems is based on the mathematics of Markov processes. Informally, they are simply rule systems that dictate permissible transitions from one node to another and produce symbols with each transition. If you imagine a series of "moves" from one node to the next in either of the diagrams and imagine that each such move allows you to copy the letter that corresponds to that transition until we exit the system in the last node, you can see how a "grammatical" string of letters can be generated. Each such legitimate sequence created in this manner is "well-formed" in the sense that it conforms to the rules of this artificial system.

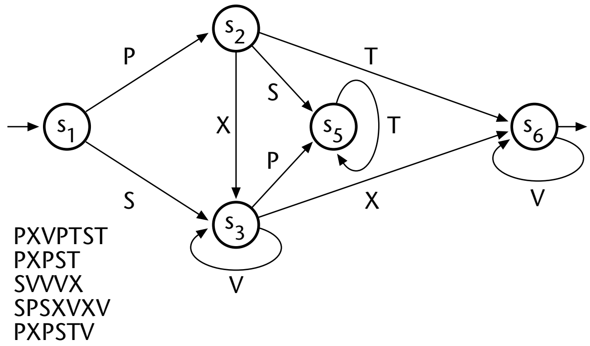

Figure 1. A sample artificial grammar with several examples of strings that it can generate using the letter set PSTVX.

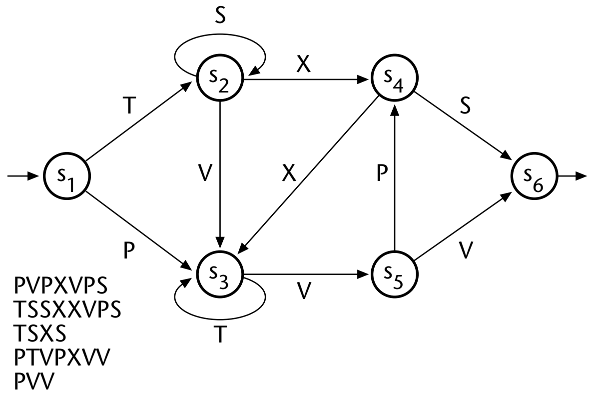

Figure 2. A second example of an artificial grammar with several well-formed strings using the same letter set.

Subjects begin these experiments thinking that they are in a simple memory experiment. It is only after the learning phase that they are informed about the existence of rules and asked to classify novel letter strings according to whether or not they are well-formed. These test strings are either new "grammatical" (that is, they are strings that they haven't seen before but are constructed following the grammatical rules) or they are "nongrammatical" (that is, they contain a letter or letters in positions that violate the rules of the grammar). For example they may be asked to judge the "well-formedness" of a string such as PTTTVPS (which, in fact, is well-formed by the AG in Figure 2) or a string like TSSXX-

VSS (which is not well-formed because it has a non-allowable transition in the next-to-last letter which can be seen by trying to generate it using the diagram).

The classic finding is that subjects are quite good at determining the grammatical status of these novel strings despite having virtually no conscious knowledge of the patterns that underlie the AG (see REBER 1993). The presumption is that subjects emerge from an experiment like this with some underlying mental representation that captures features of the regularities in the display, which is what enables them to classify items they have never seen before.

The sequential reaction time task (SRT). Here subjects sit in front of a computer monitor on which several small boxes are arranged along the bottom of the screen. A light or symbol (such as an * but it can be anything) flashes in one of the boxes and the subject is asked to press the button or key that corresponds spatially with that location as quickly as possible. On each of several hundred (or even thousand in some studies) trials a different box lights up and each time the participants must react to it as quickly and accurately as possible. In the canonical experiment the sequence of lights follows some repeating pattern or, in the study we report on here, complex rule-governed patterns that are dictated by an AG like those described above.

The classic finding is that subjects' reaction times (RT's) get faster and faster over trial blocks as they learn to exploit the structure in the sequence (see HSIAO/REBER 1998 for a review). To make sure that subjects are really learning the sequence and not merely learning a simple sensorimotor task, a random block of trials is introduced and any changes in the overall RT's to the new, nonstructured series of lights is observed. The standard finding is that RT's slow down dramatically as soon as the sequence is removed and speed up again if it is later restored. Like the AGL experiments, subjects are typically unaware of even the existence of the structured sequence let alone have access to reportable knowledge about its structure.

More on both of these procedures later.

Consolidation

In all of these situations, the knowledge, whether it is perceptual, motoric, episodic or cognitive, implicit or explicit, must, in order to be used effectively by its holder, be stored in a relatively permanent state. There have been explorations of the time course and speculations about the underlying biological changes that characterize this process whereby representations become fixed and a brief overview of them will help set the stage for our research.

Consolidation is a process whereby a memorial representation that was established during learning continues to change, after learning is complete (MCGAUGH 2000; MÜLLER/PILZECKER 1900). The process can be described both on the neuronal, biological levels (BRASHERS-KRUG/SHADMEHR/BIZZI 1996; KANDEL/SQUIRE 2000) and the psychological, behavioral levels, although it is the psychological aspect that we will focus on. During consolidation changes take place in the synapses of neurons and these changes continue to occur well after the learning procedure had been completed (KANDEL/SQUIRE 2000). These long-term synaptic changes have been shown to occur in mice, rats and other species (SQUIRE 1992). The consequences of these biological processes can be seen if they are disrupted in any of a number of different ways; for example by administering an electric convulsive shock (ECS) across the brain, or by having the subjects learn something new after learning the original material. While there are good reasons for suspecting that these procedures have rather dissimilar neurological effects, the behavioral consequences are similar in that the interference has, in some fashion, compromised the processes whereby the knowledge becomes fixed or consolidated. Much of the original work using such techniques suggests that, at least in rats and mice, five hours is usually enough to establish a long term memory trace, such that ECS will no longer interfere with the memory for a radial arm maze (see KANDEL/SQUIRE 2000 for an overview of this research).

In humans, memory consolidation seems to be dependent on structures in the medial temporal lobes (MTL), specifically the hippocampus (SQUIRE/KNOWLTON 1995). These structures are best thought of as pathways that are responsible for the passage of memory from short-term storage to long-term storage. When the hippocampus and its associated areas are damaged or destroyed it becomes difficult, and in extreme cases impossible, to form new explicit memories. The discovery of the critical role that the MTL structures play here were made in the case of HM, the first neurological patient to have his hippocampus surgically removed (MILNER 1962; MILNER/CORKIN/TEUBER 1968). HM suffered from severe and intractable epilepsy, the neural focal point of which was in the MTL. To prevent the multiple, daily seizures he suffered from, surgeons extirpated bilaterally the affected regions of his medial temporal lobes.

While the surgery was successful in stopping the seizures, HM emerged from the procedure with severe, chronic anterograde amnesia, which rendered him incapable of forming new memories. The standard interpretation of HM, and the now rather large number of patients with similar neurological insults (see SQUIRE 1992 for a review), is that they do not suffer from a learning deficit, per se, but rather from an inability to consolidate new explicit, or declarative knowledge. Patients with MTL damage show no diminished ability to recall episodes that occurred prior to the trauma, they present a nearly normal short-term memory profile, and, most importantly from our perspective, show relatively intact *implicit* learning and memory (KNOWLTON/SQUIRE 1994, 1996).

This pattern of findings suggests strongly that the underlying neurological pathways that are critical for the consolidation of explicit, declarative knowledge are distinct from those that mediate implicit, unconscious knowledge. And, indeed, there are a number of converging lines of evidence to support this distinction. In recent studies, the performance of anterograde amnesiacs has been shown to be statistically indistinguishable from that of normals on a variety of tasks including fragment completion (TULVING/HAYMAN/MACDONALD 1991), artificial grammar learning (KNOWLTON/RAMUS/SQUIRE 1992; KNOWLTON/SQUIRE 1994), and dot pattern categorization (KNOWLTON/SQUIRE 1994).

There are at least two viable hypotheses why these patient populations display these residual implicit functions. One is that implicit, nondeclarative knowledge does not need to consolidate. If knowledge that is picked up and represented implicitly were completely formed at the moment of its encoding, then we would expect to see intact performance on these tasks even in the absence of neural pathways that have been implicated in the consolidation of explicit, declarative knowledge. Tentative support for this hypothesis seemed to come from studies with intractably depressed patients who received electro-convulsive shock therapy (ECT). While ECT is known to produce both anterograde and retrograde amnesia (DUNCAN 1949; SQUIRE 1992), these patients typically display normal levels of performance on implicit tasks (SQUIRE/COHEN/NADEL 1984; VAKIL et al. 2000). Unfortunately, these studies were not designed to explore consolidation issues and little or no attempt was made to control for the time interval between the learning and the administration of the ECT. In several cases there was a full 24-hours between initial presentation and the shock therapy. Since intervals like this are known to be sufficient for consolidation to have taken place, interpreting these findings is problematical.

The other hypothesis, which we feel is, prima facie, more plausible, is that implicit knowledge, like explicit, does need to consolidate, but that the consolidation process takes place using cortical structures and pathways other than those in the MTL. We have recently begun a series of experiments designed to look more closely at this hypothesis, specifically by exploring the time course of consolidation of implicitly acquired knowledge and comparing it with what is known about other forms of memory representations. This work is carried out within a theoretical framework that builds on some basic principles of evolutionary biology that makes some fairly clear predictions about the kinds of dissociations we anticipate finding between the implicit "bottom-up" systems and the more explicit "top-down" systems that are modulated by conscious planning and control.

Evolutionary Considerations

The theory, details of which can be found elsewhere (REBER 1992a, 1992b, 1993), treats evolution as a hierarchical process in which evolutionarily old adaptive traits and structures form the foundation for the later emergence of newer traits and structures. This pyramidal process has certain consequences in terms of how characteristics of considerable evolutionary antiquity should function relative to more recently evolved ones. Since it seems fairly clear that the variety of consciousness that we are alluding to in the above discussions of explicit or declarative knowledge is unique to our species and, hence, must have evolved rather recently, we would expect the older, bottom-up implicit systems to display particular properties that the newer, top-down explicit systems do not.

In particular, the model predicts that older functions will be more *robust* than newer with those of greater evolutionary antiquity being more likely to withstand the effects of neurological and psychological disorders and injuries. As noted above, implicit functions remain largely intact in cases of amnesia. They have also been shown to withstand psychological disorders that affect explicit functions such as schizophrenia (ABRAMS/REBER 1988), Huntington's disease (KNOWLTON/SQUIRE/BUTTERS

1994), depression, and high levels of anxiety (RATHUS et al. 1994). They appear to develop normally in childhood when various syndromes compromise consciously controlled functions as in such disorders as Williams syndrome (DON et al. 2002) and autism (SMITH/REBER, in prep). In addition, there are reasons for suspecting that they should also produce memories that are particularly robust and withstand the normal, EBBINGHAUSIAN (EBBINGHAUS 1885) degradation of representation that occurs over time. Taken together these findings support the argument that implicit functions are evolutionarily and neuroanatomically distinct from explicit functions. However, they are agnostic on the question of consolidation and its time course.

Does Implicit Knowledge Consolidate and, if so, How Long Does It Take?

Despite having very specific predictions about the robustness of implicit processes, we really have no a priori hypotheses about whether cognitive representations of the type that are acquired in an implicit learning task undergo the kinds of consolidation processes that are readily seen with explicit, consciously encoded and stored memories. However, there are some hints that can be gained from looking at the parallel work carried out on two closely allied topics, implicit memory and motor learning.

Implicit memory. This issue seems like a natural one to begin with since participants emerge from an implicit learning procedure with implicit memories. However, as we noted above, there are problems with drawing too strong a parallel. One difficulty is that virtually all of the studies of implicit memory examine episodic memory for well-known material and not the traces laid down by newly acquired knowledge. This point is subtle but important. In most implicit memory experiments subjects are primed with material that they already knew (e.g., a list of familiar words, pictures of common objects) and are then later asked to carry out a task such as stem completion or degraded-picture identification. The memory here looks like it is memory for words or pictures, but the memory is also, in a functional sense, for a particular episode in which those words or pictures were used—namely this experiment with these materials run in this particular setting.

Nevertheless, the few studies that have looked at the robustness and longevity of implicit memory are relevant, although the results have been somewhat inconsistent. In a series of experiments SLOMAN et al., (1988) examined the duration of implicit memory for primed material. They found, perhaps not surprisingly, that for "ordinarily" presented materials subjects show an initial rapid forgetting within a mere 5-minute period and a continued, but much slower, loss of knowledge over the next seven days. This pattern of rapid initial loss followed by a gradual drop in performance is similar to what is typically found in studies for explicitly encoded material. However, for memories that survived the period of loss, there was evidence for rather remarkable robustness. In one study, after 16 months subjects still showed significant retention based on rapid and accurate stem-completion performance. And, of particular interest to our concerns with consolidation, SLOMAN et al. also reported that retention performance was unaffected by the learning of a second list of words even when the interpolated list was presented within 20 minutes of the primary one.

This last result, along with similar findings (GRAF/SCHACTER 1987; JACOBY 1983), prompted some to suggest that implicit representations may be immune to interference (see ANDERSON/NEELY 1996 for a review)—a conclusion that would have serious implications for any theory of consolidation. However, there are reasons for suspecting that things are a bit more complex. First, it needs to be appreciated that the representations established in these experiments are not necessarily for "just" the words used in these studies, they are likely for the full episode. If the episodic representations are deeply encoded, then subsequent learning might not produce the expected retroactive interference. There is support for this interpretation from a study by SQUIRE/SHIMAMURA/GRAF (1987) who showed a levels-of-processing effect in which words encoded semantically showed better and longer-lasting priming effects than those processed nonsemantically. Interestingly, this study also compared the performance of amnesic patients with MTL damage with a group of matched normal participants. The amnestic group showed very similar priming effects—but only when retesting was carried out within two hours and nonsemantic encoding was encouraged. Normals showed highly robust priming effects after a lag of four days and a significant boost in performance when deeper levels of processing were encouraged.

Recently, LUSTIG/HASHER (2001a, 2001b) have presented convincing arguments and data that show that the nature of the encoding in these experiments is critical. They suggest that the lack of interference observed in the earlier studies was the result of using interpolated word lists that were not sufficiently similar to the target lists. When they used distracter lists that were highly similar to the target list, they found that performance on word fragment completion was disrupted. From our perspective, this interpretation feels appropriate. Implicit memory likely undergoes some form of encoding and its ultimate representation should, in principle, be affected by interpolated tasks that share such encoding. Unfortunately for the primary focus of this investigation, LUSTIG and HASHER never manipulated the time between initial learning and the introduction of the second task so we have no hints as to what the time course of consolidation might look like under these conditions.

Motor learning. With regard to implicit procedural learning the effects of interfering tasks have been studied methodically and with greater attention to the question of the time course of consolidation (BRASHERS-KRUG/SHADMEHR/BIZZI 1996). However, the studies here have all been carried out using a single procedure. Consequently, while the results are intriguing and potentially of considerable importance, we still need to exercise caution concerning the generality of the findings.

In the canonical experiment subjects are trained to move a lever in a complex pattern of movements. The lever is attached to and controlled by a robot that can be programmed with a variety of force vectors that push and/or pull the lever in particular directions. The subject's task is to learn to counteract the mechanical forces of the device and learn to make the required movements smoothly and efficiently. The interpolated, interfering task appears identical to the initial task only the robot arm is reprogrammed so that the original counteractive movements are no longer effective and new ones must be learned.

With this task BRASHERS-KRUG/SHADMEHR/BIZZI (1996) found dramatic retroactive interference effects. If the second task is carried out immediately after the first, it virtually wipes out any representation of the original learning. Subjects in this condition show essentially no savings when retested on the first task. However, within a mere five minutes a consolidation effect begins to emerge. Subjects given a short rest period before the second task show a small but reliable savings effect. As the time between the two tasks is lengthened to 1 hour, 4 hours and 24 hours, a monotonic consolidation effect emerged. Interestingly, the performance of the participants in the 4-hour group did not differ from those in the 24-hour group suggesting that, under these conditions at least, 4 hours appeared to be sufficient for the motor schema to become fully consolidated. The suggestion is that immediately after learning, the implicit memorial representation of the learned domain exists in a rather fragile state and is relatively easy to disrupt. However, as time passes the memory becomes stronger, more robust, and less prone to interference.

In an important follow-up, SHADMEHR/HOLCOMB (1997) discovered particular neural correlates to this consolidation process. Subjects who practice the task again after a lag of roughly 6 hours show, as expected, no change in performance. However, PET scans revealed that the brain now engages new regions when performing the task. During the initial performance the most active brain areas were in the prefrontal cortex. Six hours later, after consolidation, they observed a shift from these regions to premotor, posterior parietal and, perhaps not surprisingly, cerebellar cortex.

These findings are tantalizing for a number of reasons. For one they support the argument that the consolidation process is accompanied by a shift in the underlying neural structures that mediate the behavior. Second, they suggest strongly that the establishment of lasting memorial representations for a sensorimotor task takes place without calling upon structures in the medial temporal lobe, which are known to be essential for the establishment of explicit memories. Third, they imply that the time course of consolidation of motor learning is a monotonic function with an asymptote around five or six hours, suggesting that sleep is not an integral element. This latter point is rather intriguing since, for nearly 80 years now (JENKINS/DALLENBACH 1924), the dominant view of the consolidation of tasks with a cognitive flavor is that sleep (either REM or slow wave or perhaps both) is required.

Thus, implicit memory and implicit motor learning share the properties of being easily interfered with immediately after learning when the interpolated material is similar in nature to the initial representation. However nothing is known about whether a more cognitively oriented implicit learning task is prone to this kind of interference or whether it shows a similar time course. With these thoughts in mind, we began a series of studies into the time course of consolidation of implicitly acquired knowledge. While the work is still very much in progress we feel sufficiently confident in the early findings to present some of them here with our best guess as to how the others will turn out.

Some Pilot Data

In the initial experiment we examined implicit learning by exposing people to a complex, ordered environment and monitoring their ability to learn the regularities of that environment. We explored the consolidation of such learning by interposing a second task designed to produce maximum interference. This study used the artificial grammar learning (AGL) task described earlier.

Initially we asked our subjects to memorize 20 grammatical strings from one of the two AG's. We then tested the knowledge for the underlying rules of the grammar by presenting them with one hundred novel letter strings, fifty of which were well-formed and fifty of which are not and asking them to classify each according to their sense of its grammaticality. Our group of participants correctly classified an average of 63% of the test strings, demonstrating that they have indeed learned a considerable amount about the underlying structure of the grammar. These results are fairly typical of AGL tasks (see REBER/ALLEN/REBER 1999 for a review).

However, the question that interested us was whether this knowledge, once established, was vulnerable to interference and, if so, how long would it be in this unstable state? We introduced an interpolated task using the other AG shown in Figure 2. Note that the two AG's were instantiated using the same five letters thereby producing letter sequences that look superficially very much like each other. We took seriously the message in LUSTIG and HASHER's (2001a, 2001b) studies that such similarity is critical for disrupting the memory traces of initial learning.[1]

We ran three groups each with a different delay between the first grammar and the second: 5-minutes, 5-hours, and 24-hours. The first delay condition allows us to see what kind of interference we get immediately after learning. The second matches the delay that SHADMEHR and his colleagues found was sufficient for consolidation of a complex sensorimotor skill. The third was introduced to explore the tantalizing possibility that this task, which is considerably more cognitive in nature, undergoes additional consolidation during sleep.

The procedure was straightforward. After subjects were tested for knowledge of the first AG, they were run through a second memorization phase using letter strings from the other. Immediately following the learning of the second grammar, subjects were asked to go through the well-formedness test again for the test strings from the first AG. Knowledge for the second AG was not evaluated. The key data are the subjects' scores on the well-formedness task on the second run compared with how well they did first time through.

The results were quite striking. The 5-minute and the 5-hour condition produced very large interference effects compared to the 24-hour condition. In fact, subjects' ability to make well-formedness judgments in the first two groups dropped from 63% to levels that were barely above chance. It was almost as though the second AG "erased" any residual representations of the first. Note that this result was not a matter of simple memory decay. Matched control groups who were retested on the first AG using the same 5-minute and 5-hour delays but without any specific interpolated experience showed virtually no drop-off in performance. The 24-hour group, however, showed significantly greater resistance to interference. While their scores were slightly below those of their control group, they retained their ability to reliably make well-formedness judgments.[2]

These results are intriguing and, to some extent, surprising. We found no evidence for the monotonic increase in consolidation that SHADMEHR and his colleagues reported. In fact, we failed to see any significant savings unless participants were allowed to return the next day for the interpolated task. One possible interpretation is that our task, having much more in the way of a "cognitive" flavor to it than the more sensorimotor task that SHADMEHR and coworkers traditionally use, requires a period of sleep in order to consolidate.

We have just begun an investigation into this latter possibility. We are using a sequential reaction time (SRT) task that is structurally similar to the AG task but where the actual behaviors are expressed in a manner that is generally regarded as having a fundamental and critical motor component (see WILLINGHAM 1992; WILLINGHAM et al. 2000). As described above, in the SRT task participants sit in front of a computer screen with five locations marked across the bottom of the screen. What we did was to have each location correspond to one of the letters used in the AG task. Rather than showing subjects sequences of letters generated by an AG and asking them to memorize them, we presented them with sequences of lights in particular locations where the order of locations followed an AG. For example, rather than seeing a letter string like PTTTVPS, they saw a light flash in position 1 followed by 3 flashes in position 3, followed by single flashes in positions

4, 1 and 2. The sequence continues by juxtaposing an extended series of strings one after the other so that several hundred trials can be presented in a smooth, unbroken sequence. Our procedure was to run subjects through 9 blocks of such trials with 100 trials in each block. Following this learning phase, we introduced the "transfer" block in which the structured sequence of lights is replaced with a random series. The experiment ended with a final trial-block in which the grammatical sequence was re-introduced. Learning was assessed by comparing the RT's on the transfer block with the two blocks of structured sequences that bracket it.

The second, interpolated task was a 9-block experience with a randomly generated sequence. And, following the logic of the earlier experiment, this was followed by a restoration of a "grammatical" sequence for a full trial-block. Two groups have been run to date, one with a 5-minute lag between the initial learning with the structure sequence and the random and one with the critical 24-hour rest period.

With all the standard caveats about preliminary results, the data show an intriguing pattern that rather neatly parallels the findings with the AG task. The introduction of the random training five minutes after initial learning dramatically reduces performance. In fact, in this case it seemed to remove any evidence of residual learning. When the structured sequence was re-introduced, subjects RT's went back up again to a level that was not different from that observed early in training. However, in the 24-hour condition, we observed almost complete retention of the sequence. Here the interpolated task had virtually no disruptive effect. When we restored the sequence, subjects' RT's dropped immediately to levels statistically indistinguishable from those seen at the end of the initial learning phase.

Unfortunately, at this juncture we have too little data from a 5-hour group to report on—although however the results come in, they will be revealing. If we see little in the way of consolidation over this time period, it will suggest that a key difference between our procedures and those of SHADMEHR and his colleagues is the complexity of our task relative to theirs. However, should we observe significant resistance to interference from the interpolated task, it would strongly suggest that the nature of the task is central to the time course of consolidation with motor tasks undergoing the required molecular changes more rapidly than cognitive tasks and, importantly, without needing a sleep period.

The results from these two pilot studies are certainly consistent with the consolidation hypothesis originally proposed by MÜLLER/PILZECKER (1900) and later extended by many others including MILNER (1970), REBER/SQUIRE (1998), FREY/MORRIS (1997), SHADMEHR/HOLCOMB (1997), and BRASHERS-KRUG/SHADMEHR/BIZZI (1996). The data accumulated across these studies shows that after learning a period free from inputs that call for the establishment of competing representations is more conducive to long-term memory formation than a period containing such competing tasks. Consolidation appears to be a requisite for establishing relatively fixed implicit memories just as it is for declarative, explicit representations. So far as we can determine, the two studies we report on here are the first to show that the consolidation process applies to a type of learning that is largely unconscious and yields abstract representations. Moreover, since these implicit acquisitional operations are known to be intact in amnesic individuals, these findings suggest strongly that non-hippocampal structures are being recruited for this process. Several significant questions, of course, remain.

General Discussion

■ First, how important is the type of learning? Does the knowledge of a structured display acquired via a cognitive task like AG learning consolidate in the same manner as it does for a more motoric task like the SRT procedure? Our preliminary data suggest that representations based on these two may differ in both the robustness of the representations and their time courses. We hope to have more to say about this problem in the near future.

■ Second, how important is a period of sleep in the consolidation process—and does the type of task interact here? Again, the issue is far from resolved. SHADMEHR and his colleagues found virtually complete consolidation without sleep. Our data suggest that, as least for the more cognitively oriented AG task, sleep is important. However, as before, the information contained in these tasks and the type of learning are confounded.

■ Third, if there are no interfering tasks interposed, does performance increase with consolidation? In SHADMEHR/HOLCOMB's (1997) study, performance actually increased in conditions without the interfering task. This was particularly true for subjects who were retested 24 hours later. However, in our studies we saw little evidence for such improvement. In the AG task we observed no increase in performance over a 24-hour period and in the SRT task it is exceedingly difficult to make this kind of comparison. Very little is known about the differential effects that sleep might have on the formation of explicit, implicit cognitive/perceptual, and implicit sensory/motor knowledge. Research along these lines might prove to be interesting.

Finally we'll close on an "up" note. We are hopefully beginning a long series of investigations that should allow us to unpack the complexities hinted at here. So far we see no reason for suspecting that implicitly acquired and represented knowledge does not undergo a consolidation process that has basic principles of molecular biology at its core. What we don't know, of course, is whether these molecular processes share features and properties with other, better known mechanisms. We also need to pursue the question of cortical locus, how does such a mechanism operate independent of the hippocampal structures known to function with declarative, explicit knowledge? And, finally, we will need to unpack the role played in these processes by the type of learning. Do perceptual, cognitive and motor learning mechanisms produce epistemic contents that consolidate in the same manner or does the mode of acquisition have an effect on how and how rapidly the representations become robust? The future looks very interesting.

Authors' address

Leib Litman and Arthur S. Reber, Department of Psychology, Brooklyn College of CUNY, Brooklyn, NY, 11210, USA.
Emails: psychdrum@yahoo.com, areber@brooklyn.cuny.edu.

Acknowledgement

Preparation was supported by a grant from the PSC-CUNY and Grant #0113025 from the National Science Foundation.

Notes

1 The AG's were counterbalanced so that half the subjects learned with letter strings generated by one grammar with the other used for producing the interpolated task and the reverse order was used with the remaining participants.

2 In fact, their relative drop-off in performance from the first, base-line testing was roughly the same as that reported by ALLEN/REBER (1980) where the second AG learning task was introduced a month after the first (lots of time for consolidation!) and the restesting for knowledge of the first AG was carried out a full two years later. While cross-experiment conclusions are always made with caution, this pattern of results does suggest that once knowledge such as this has consolidated it may be permanent

References

Abrams, M./Reber, A. S. (1988) Implicit learning: Robustness in the face of psychological disorders. Journal of Psycholinguistic Research 17:425–439.

Allen, R./Reber, S. R. (1980) Very long term memory for tacit knowledge. Cognition 8:175–185.

Anderson, M. C./Neely, J. H. (1996) Interference and inhibition in memory retrieval. In: Bjork, E. L./Bjork, R.A. (ed) Memory. Academic Press: San Diego, pp. 237–313.

Berry, D. C./Broadbent, D, E. (1984) On the relationship between task performance and associated verbalizable knowledge. Quarterly Journal of Experimental Psychology 36:209–231

Brashers-Krug, T./Shadmehr, R./Bizzi, E. (1996) Consolidation in human motor memory. Nature 382:252–255.

Duncan, C. P. (1949) The retroactive effect of electroshock on learning. Journal of Comparative and Physiological Psychology 42:32–44.

Don, A. J./Schellenberg, E. G./Reber, A. S./DiGirolamo, K. M./Wang, P. P. (2002) Implicit learning in children and adults with Williams Syndrome. Developmental Neuropsychology. In Press.

Ebbinghaus, H. (1885) Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Duncker & Humblot: Leipzig.

Frey, U./Morris, R. G. (1997) Synaptic tagging and long term potentiation. Nature 385:533–536.

Graf, P./Schacter, D. L. (1987) Selective effects of interference on implicit and explicit memory for new associations. Journal of Experimental Psychology: Learning, Memory and Cognition 13:45–53.

Hsiao, A. T./Reber, A. S. (1998) The role of attention in implicit sequence learning: Exploring the limits of the cognitive unconscious. In: Stadler, M. A./Frensch, P. A. (eds) Handbook of implicit learning. Sage: Thousand Oaks CA, pp. 471–494.

Jacoby, L. L. (1983) Perceptual enhancement: Persistent effects of an experience. Journal of Experimental Psychology: Learning, Memory and Cognition 9:31–48.

Jenkins, J. G./Dallenbach, K. M. (1924) Obliviscence during sleep and waking. American Journal of Psychology 35:605–612.

Kandel, E, R./Squire, L, R. (2000) Breaking down scientific barriers to the study of brain and mind. Science 10(290):1113–1120.

Knowlton, B, J./Ramus, S, J./Squire, L, R. (1992) Intact artificial grammar gearning in amnesia: Dissociation of Classification Learning and explicit memory for specific instances. Psychological Science 3:172–179.

Knowlton, B. J./Squire, L, R. (1994) The information Acquired in artificial grammar learning. Journal of experimental Psychology: Learning, Memory and Cognition 20:79–91.

Knowlton, B. J./Squire, L. R./Butters, N. (1994) Intact artificial grammar learning in patients with Huntington's disease. Society for Neuroscience Abstracts 20:1075.

Lewicki, P. (1986) Nonconscious social information processing. Academic Press: New York.

Lustig, L./Hasher, L. (2001a) Implicit memory is vulnerable to proactive interference. Psychological Science 12(5):408–412.

Lustig, L./Hasher, L. (2001b) Implicit memory is not immune to interference. Psychological Bulletin 127:629–650.

McGaugh, J. (2000) Memory: A century of consolidation. Science 287:248–251.

Milner, B. (1962) Les troubles de la mémoire accompagnant des lésions hippocampiques bilatérales [Disorders of memory accompanying bilateral hippocampal lesions]. In: Pysiologie de l'hippocampe. Centre National de la Recherche Scientifique: Paris, pp. 257–272.

Milner, B./Corkin, S./Tueber, H.L. (1968) Further analysis of the hyppocampal amnesiac syndrome. Fourteen year follow up of H. M. Neuropsychologia 6:215–234.

Milner, B. (1970) Memory and the temporal region of the brain. Academic Press: New York.

Müller, G. E./ Pilzecker, A. (1900) Experimentelle Beiträge zur Lehre vom Gedächtnis. Zeitschrift für Psychologie 1:1–300.

Rathus, J. H./Reber, A. S./Manza, L./Kushner, H. M. (1994) Implicit learning and affective states. Perceptual and Motor Skills 7:163–184

Reber, A, S. (1976) Implicit learning of synthetic languages: The role of instructional set. Journal of Experimental Psychology: Human Learning and Memory 2:88–94.

Reber, A, S.(1992a) An evolutionary context for the cognitive unconscious. Philosophical Psychology 5:33–51.

Reber, A, S. (1992b) The cognitive unconscious: An evolutionary perspective. Consiousness and Cognition 1:93–133.

Reber, A. S. (1993) Implicit learning and the cognitive unconscious. Oxford Univ. Press, New York.

Reber A. S./Allen, R./Reber, P. J. (1999) Implicit and explicit learning. In R. Sternberg (Ed.), The nature of cognition, Cambridge, MA: MIT Press. pp475–514.

Reber, A. S./Kassin, S. M./Lewis, S./Cantor, G. W. (1980) On the relationship between implicit learning of a complex rule structure. Journal of Experimental Psychology: Human Learning and Memory 6:492–502.

Reber, P. J./Squire, L. R. (1998) Encapsulation of implicit and explicit memory in sequence learning. Journal of Cognitive Neuroscience 10:248–263.

Schacter, D. L. (2001) Forgotten ideas, neglected pioneers: Richard Semon and the story of memory. Psychology Press: Cambridge MA.

Shadmehr, R./Holcolm, H. H. (1997) Neural correlates of motor memory consolidation. Science 277(8):821–824.

Sloman S. A./Gordon Hayman, C. A./Ohta, N./Law, J./Tulving, E. (1988) Forgetting in primed fragment completion. Journal of Experimental Psychology 14(2):223–239.

Smith, C./Reber, A. S. (in prep) Only nonsocial implicit learning is intact in autism.

Squire, L, R. (1992) Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. Psychological Review 2:195–231.

Squire, L. R./Cohen, N.J./Nadel, L. (1984) The medial temporal region and memory consolidation: A new hypothesis. In: Weingartner H./Parker, E. S. (eds) Memory consolidation: psychobiology of cognition. Lawrence Erlbaum Associates: Hillsdale NJ, pp. 185–210.

Squire, L. R./Knowlton, B. J. (1995) Memory, hippocampus, and brain systems. In: Gazzaniga, M. S. (ed) The cognitive neurosciences. MIT Press: Cambridge MA, pp. 825–837.

Squire, L. R./Shimamura, A. P./Graf, P.(1987) Strength and duration of priming effects in normal subjects and amnesic patients. Neuropsychologia 25:195–210.

Stadler, M. A./Frensch, P. A. (ed) (1998) Handbook of implicit learning. Sage: Thousand Oaks CA.

Tulving, E. (1983) Elements of episodic memory. The Clarenton Press: Oxford.

Tulving, E./Hayman, A. C./MacDonald (1991) Long-lasting perceptual priming and semantic learning in amnesia: A case experiment. Journal of experimental Psychology: Learning, Memory and Cognition 17:595–617.

Vakil, E./Grunhaus, L./Nagar, I./Ben-Chaim, E./Dolberg, O. T./Dannon./Schreiber, S. (2000) The effect of electroconvulsive therapy (ECT) on implicit memory: Skill learning and perceptual priming in patients with major depression. Neuropsychologia 38:1405–1414.

Willingham, D. B. (1992) Implicit motor sequence learning is not purely perceptual. Memory and Cognition 27:561–572.

Willingham, B. D./Wells, A./Laure, B./Farrel, M. J./Stemwedel, E. M. (2000) Implicit motor sequence learning is represented in response locations. Memory and Cognition 28(3):366–375.

A. Minh Nguyen

Blindsight and Unconsciousness

Abstract

This paper argues that blindsight phenomena involve unconscious experiences. The argument proceeds in two steps. First, some blindsight subjects have some of the visual experiences they deny having. Second, these experiences differ so much from normal conscious visual experiences that they must be unconscious. The paper also considers four objections against the main contention. It is argued that none of these works.

Key words

Blindsight, consciousness, unconscious experience.

Blindsight subjects are patients with visual field defects caused by occipital lobe damage. According to Weiskrantz (1986, p166), blindsight subjects "can in fact detect and discriminate certain types of visual events within their 'blind' fields by pointing or, less well, by eye fixation without, however, being aware that they can do so: they think they are only "guessing'". Weiskrantz goes on to define "blindsight", more compactly, as "visual capacity in a field defect in the absence of acknowledged awareness".

I believe that the vast range of blindsight phenomena constitutes a substantial body of evidence for the existence of unconscious experiences. By "unconscious experiences" I just mean those that are not conscious. Most philosophers take it for granted that there are no such things as unconscious experiences. My paper aims to show, however, that this widespread assumption is probably false.

To show that blindsight phenomena involve unconscious experiences, it suffices to show that some blindsight subjects have some of the visual experiences they deny having, and that these experiences differ so much from normal conscious visual experiences that they must be unconscious. By "normal conscious visual experiences" I mean those that can be found in most if not all cases of normal vision.

Now perhaps the first thing that must be noticed about a blindsight subject is that, despite her denial that she sees anything at all within her blind field and despite her insistence that she just guesses all the way through, her performance in a forced-choice detection or discrimination task is typically well above chance. Such performance reaches a perfect or near-perfect level in some cases—for example, movement and orientation discriminations. It is reasonable to conclude, in light of the general reliability of her avowed guesses, that the blindsight subject is in some states which bear matching information about certain stimuli in her impaired field. These states cause and explain, or help cause and explain, her usually correct (albeit forced) choices. That is: while making the discriminations and detections that she does, the blindsight subject behaves as though she occupied certain information-carrying states of some kind; it is reasonable to explain her apparently informed actions by citing these states as their causes, or parts of their causes.

So far as I know, the above information-carrying states are either sensory or intentional. Sensory states such as pain and sensation of color have qualitative or experiential aspects but lack propositional or conceptual contents, whereas intentional states such as belief and desire have propositional contents but lack qualitative aspects. Qualitative aspects are the ways things seem to us. If you look at a computer screen, the way it looks to you—the subjective visual quality of the computer screen—is the qualitative aspect of your visual experience. Propositional contents, in contrast, are what are expressed by the sentences that occur in the that-clauses, the ones that are embedded in intentional-state ascriptions. In the ascription "Nat believes that the cat is on the mat", the sentence that follows the "that" expresses the propositional content of the belief that is ascribed to Nat.

I do not mean to suggest that qualitative aspects and propositional contents cannot cohabit with each other. Emotions such as sadness and joy are ordinarily characterized both in terms of how things feel to the person having them and in terms of the things his feelings are directed at. But the information-carrying states that the blindsight subject occupies are obviously not emotional states. Nor are they instances of any other type of mental state that exhibits this kind of duality. At least there is no reason to assume that the opposite is the case.

Now if, on the one hand, the above information-carrying states are sensory, then the blindsight subject has some visual experiences of some of the objects he says he cannot see and the first step of our argument is established. After all, sensory states are states that have distinctive sensory qualities. States that have distinctive sensory qualities are sensations. And sensations are experiences. What is more, since the information-carrying states that the blindsight subject occupies bear matching information about certain stimuli in his impaired field, the visual experiences that he thus has are therefore veridical. For veridical experiences are those whose informational contents match the relevant parts of what go on in the subject's perceptual fields.[1] Veridical experiences stand in contrast with unreal experiences such as those associated with hallucination, illusion, imagination, and dream.

If, on the other hand, the above information-carrying states are intentional, then the blindsight subject has some quasi-doxastic or belief-like states with true propositional contents about some of the objects he says he cannot see. True, the blindsight subject is not confident enough to believe what he says he only guesses. But the apparent educatedness, as is evidenced by the remarkable accuracy, of his avowed guesses would be miraculous in the absence of any cognitive stand on his part, however racked with doubt, impoverished in content, and short-lived the stand may be. Perhaps quasi-doxastic states (more precisely, their propositional contents) are largely unavailable as premises in practical deliberation and theoretical reasoning, and largely unavailable for (rational) control of speech and non-linguistic action. If this is the case, then they most likely fail to exhibit what BLOCK (1995) calls "access consciousness". Perhaps, in realizing (if ever) such an inferential or a regulative role, quasi-doxastic states are largely uninfluential. Perhaps, though they are not beliefs themselves, the intentional states that serve to record and store impressively accurate visual information about objects in the blindsight subject's damaged field would, under suitable conditions, play a role in the proximate causal history of beliefs about these objects by functioning as causal antecedents of the beliefs in question. If quasi-doxastic states play such a role, then they are "subdoxastic states" in STICH's (1978) sense. But whether these speculations on the nature of quasi-doxastic state prove true or not is independent of the following claim of mine: namely, that the fact that the blindsight subject instantiates, or is capable of instantiating, a wide range of quasi-doxastic states with true propositional contents about (the presence, absence, location, flux, movement, motion direction, orientation, shape, size, and perhaps even color[2] of) objects within his blind field indicates that he instantiates, or is capable of instantiating, at least some veridical visual experiences of these objects.

There may be representational systems (e.g., instances of certain types of machine) that cannot undergo any experience, but somehow still manage to acquire conceptually and linguistically encoded bits of information that they can then use to regulate and control their behavior. Conversely, there may be creatures (e.g., members of certain animal species) that are experiencers, but lack the conceptual resources required for the formation of propositional attitudes. But, to the extent that an individual is capable of undergoing all sorts of perceptual experience, visual experience included, and a blindsighter is such an individual, she cannot have a great variety of true quasi-doxastic states about items within her blind field without having at least some veridical visual experiences of some of these items, experiences that service the construction of some of these quasi-doxastic states. For, to produce a huge set of cognitive states about items within her blind field, the blindsighter *qua* experiencer-cum-conceptualizer cannot help drawing on sensorily encoded bits of information contained in some non-conceptual representations (i.e., experiences) of hers, and part of the best explanation of most such states' being true is that some of the utilized representations are veridical visual experiences of some of the aforementioned items. Since a conceptually competent and appropriately attentive human being who has a (veridical) experience of a certain thing tends to have some (true) cognitive state about that thing, the fact that a blindsighter has a huge set of true quasi-doxastic states about certain objects within her blind field can be properly, if partly, explained by the hypothesis that she has some veridical visual experiences of some of these objects. Thus, it is highly probable that a blindsighter has some veridical visual experiences of some objects within her blind field. The apparently legitimate presumption in favor of the truth of her repeated avowals that she does not see anything within her blind field can under some circumstances be overturned by behavioral evidence plus theoretical considerations to the contrary.

Having shown that some blindsight subjects have some of the visual experiences they deny having, let us now show that the experiences they possess in virtue of blindsighting are unconscious. To my knowledge, nobody has claimed that blindsighters still enjoy conscious visual experiences of objects within their blind fields. For to do so would amount to making the implausible claim that a human subject's capacity for conscious vision is in no way diminished by the surgical removal of his entire striate or primary visual cortex, the region of the brain to which the main visual pathway projects from the retina, and at which incoming signals from the eye ("visual inputs", in WEISKRANTZ's parlance) are normally first received or gathered. It seems universally, and rightly, acknowledged that, insofar as blindsight subjects have visual experiences of objects within their blind fields, these experiences are most likely unconscious.

But why should this be so acknowledged? We should ascribe unconsciousness to blindsight subjects' visual experiences of objects within their blind fields because, unlike normal conscious visual experiences, these experiences are relatively impoverished in content, and play no comparably robust role in effecting speech, non-linguistic action, or mental operation. The visual experiences found in blindsight differ so much from normal conscious visual experiences that they must be unconscious. The difference consists in the fact that (1) they are relatively impoverished in content, (2) their possessors not only fail to acknowledge but also deny having them, (3) their possessors are relatively incapable of initiating the kind of action that exhibits flexibility and creativity, and (4) their possessors fail to form perceptual beliefs about those stimuli in their blind fields that cause the visual experiences in question. Thus, there are four considerations for thinking that the visual experiences found in blindsight differ so much from normal conscious visual experiences that they must be unconscious. Perhaps none of these points to a difference that matters. But a combination of all four probably does. I shall examine each consideration in turn.

The first consideration is this. Conscious visual experiences are normally blessed with informational abundance and specificity. To be convinced of this, a skeptic possessed of normal eyesight needs only to look attentively at some middle-sized objects that are sufficiently near, relatively stationary, in full view, and in good light. On the whole, conscious experiences "are so rich in content—so full of information about the perceived world—that one has the sense that no practical description or catalog could do justice to them" (DENNETT 1995, p252). By contrast, a blindsight subject's residual vision of what goes on inside her scotoma is typically indistinct and color-free. It is so degraded that it has vanishingly little content. "Relative poverty of content is a non-optional hallmark of blindsight" (DENNETT 1995, p253).

The second consideration is this. Conscious experiences are such that individuals who have them tend to assert that they have them, not that they do not have them, given that they understand the question as to whether they have the experiences under consideration, possess the conceptual and linguistic resources to construct an intelligible answer to the above question, want to answer this question honestly, harbor no overriding or conflicting desires, and so on (ROSENTHAL 2003). But, in spite of the satisfaction of all these conditions and blindsight subjects' well-above-chance performance in the relevant forced-choice guessing games—the latter of which, as I have argued, indicates that some visual capacity is exercised and some visual mechanism is at work— blindsight subjects invariably, and incorrectly, insist that they do not see anything at all in the blind regions of their visual fields. It is worth noting that DENNETT seems to put the two considerations just described together by conjecturing that it is precisely because certain visual experiences of blindsight subjects' suffer from poverty of content that the subjects normally, and falsely, deny having them. This conjecture is not wildly implausible. For consider a very rough analogy: it is precisely because certain material objects are so tiny and, shall we say, perceptually unimpressive that, when asked, we typically deny their existence even though they are there.

The third consideration is this. Individuals possessed of conscious experiences tend to initiate rational or self-benefiting actions with respect to those parts of the world that they consciously experience or perceive. For instance, those gazelles that consciously see, hear, or smell the lions tend to take flight, and those people who not only have but also feel pain tend to seek relief. But this is not true of visual experiences found in the case of blindsight, for the contribution they make to the causal process leading to actions is relatively limited. Blindsight patients are much less capable of making fine discriminations and much less flexible than their normally sighted counterparts. Such patients must be cued to give their better-than-chance guesses, and it is doubtful that, without prompting, the patients' visual experiences of objects within their blind fields still tend to result in actions, especially rational actions. As DRETSKE (1997a, p13) puts it, "people afflicted with [dissociation] syndromes are always 'deeply disabled'… Blindsight subjects cannot avoid bumping into lampposts, even if they can guess their presence or absence in a forced-choice situation. Furthermore, '[a]ll these subjects lack the ability to think about or to image the objects that they can respond to in another mode, or to inter-relate them in space and in time; and this deficiency can be crippling.'"

DRETSKE's last quoted remark already slides us into the fourth, and final, consideration, which is this. A conscious experience of a particular object typically if not invariably produces in a subject who not only undergoes the experience but also masters the germane concepts a belief about that object. For it is impossible, in my view, to conceive that a creature consciously senses a certain object, grasps the appropriate concepts, thinks that he is not dreaming, imagining, or hallucinating, and nevertheless fails to form any perceptual belief about that object. Since I consider inconceivability a reliable sign of impossibility, I conclude that, as a general rule if not an exceptionless one, conscious sensory experiences of the kind of creature described above are part of the causal process leading to perceptual belief formation. But I do not think that any visual experience issued from blindsighting is part of such belief-forming process. After all, the blindsight subject, who takes himself to be only guessing, is not self-assured enough to believe that there exists such and such object in his blind field (the content of this belief is henceforth abbreviated as *e*). If he believed that *e*, he would be ready, in suitable circumstances, to assert "*e*", draw certain inferences and perform certain actions on the assumption that *e*, but his actual behavior does not indicate such readiness.

The cognitive states to which the visual experiences connected with blindsight standardly give rise are not beliefs but merely quasi-doxastic states. Such experiences usually play a role in the (proximate) causal history of the quasi-doxastic states that underlie the blindsight subject's statistically significant performance in the relevant forced-choice situations. But they will never function as causal antecedents of beliefs unless the quasi-doxastic states produced by them transform into something like subdoxastic states. A prime example of subdoxastic states is the mental states that serve to represent the rules of grammar of the subject's language. Now "subdoxastic states, as contrasted with beliefs, *are largely inferentially isolated from the large body of inferentially integrated beliefs to which a subject has access*" (STICH 1978, p507). Hence, since quasi-doxastic states are not more cognitively advanced than subdoxastic states, they are less inferentially integrated than beliefs. This gives us still another reason for accepting the hypothesis that the visual experiences connected with blindsight play a less robust role in belief formation than conscious experiences. For the hypothesis—together with the fact that quasi-doxastic states are less inferentially integrated than beliefs and hence less effective in influencing behavior, including mental behavior—best explains why blindsight subjects, whose visual experiences of objects within their blind fields typically give rise to only quasi-doxastic states, are less capable of making fine discriminations and less flexible than their normally sighted counterparts, whose conscious visual experiences typically give rise to beliefs. And we can infer, from the fact that the hypothesis offers the most plausible explanation of the evidence, that the hypothesis is true.

I take blindsight to be intermediate between mere pupil reflex and normal conscious vision (STOERIG 1997). For mere pupil reflex involves no visual experience, while the visual experiences found in blindsight are unconscious whereas those found in normal vision are (typically) conscious. Mere pupil reflex involves just the functioning of a certain motor (as opposed to sensory) cortex. Such a mechanism has direct descending connections to motor neurons and serves to regulate muscular movements and glandular secretions (REBER 1985, p455). Our pupils reliably expand in darkness and contract in bright light. That is how they automatically and mechanically respond to signals of such and such levels of light intensity, no matter whether we are awake or sleeping, alert or comatose, sighted or blind. Similar pupil reflexes are reliably elicited by similar amounts of light falling into the eyes. These happen even if the stimuli fail to be mentally represented by the living subjects whose pupils reflexively respond to them, and even if the subjects are totally asleep, blind, or even unconscious. By contrast, both blindsight and normal vision involve visual experiences and hence "creature consciousness".[3] But while superior, in terms of mental responsiveness, to mere pupil reflex, blindsight is nonetheless inferior, in terms of "state consciousness", to normal vision. For the visual experiences found in normal vision, as a general rule, have every one of the features that we commonly assign to conscious states, whereas those found in blindsight lack them all.

This concludes the second step of my argument. To stave off one potential misunderstanding, some clarification is in order. I claim that there are four considerations for thinking that the visual experiences found in blindsight differ so much from normal conscious visual experiences that they must be unconscious. These considerations have to do with relative poverty of content, failure to make correct psychological self-reports, relative incapacity for engaging in flexible goal-directed behavior, and failure to form perceptual beliefs. Now each of these may not mean much in and of itself. But when joined with the rest it becomes quite crucial. For in such union there is real cause for state unconsciousness. Or so I have been contending. It is my contention that if an experience has every one of the four qualities mentioned above, it is unconscious. Since the visual experiences found in blindsight have all of these qualities, they are unconscious.

It is consistent with my view that some experiences have some of the above qualities and fail to be unconscious. Such experiences in fact exist. Consider conscious non-foveal vision. Our visual experiences of objects in the peripheral regions of our visual fields bear little content. They carry little information about the perceived world. But to the extent that we are aware of our own peripherally visual experiences, able to deal flexibly with objects on the periphery, and poised to form some basic beliefs about the existence or character of these objects, there is no reason to deny that the experiences in question are conscious. On the other hand, and this is one of my main points, if we lack such awareness, such ability, and such poise, it does not seem wrong to deny this.

Similarly, after-images do not carry much information. Nor do they normally cause perceptual beliefs about the sources of blazing light that have stimulated the eyes. For we either form or do not form such beliefs while our eyes are being stimulated. If we do not, then it is unlikely that we will do so thereafter. After all, the condition will be less favorable. If we do, then such beliefs must be caused by something else. After all, after-images occur only after the intense stimulation of the eyes has ceased. But despite the fact that after-images fail to carry much information and fail to cause perceptual beliefs, such intense experiences are nonetheless often conscious.

I now consider four objections to what has been said so far. The first objection is that there could be no such thing as an unconscious experience. Such an experience is a logical impossibility. Since I argue for the existence of unconscious experiences, my argument is unsound. That is, either the argument is invalid or at least one of its basic premises is false.

No doubt many people use the term "experience" to denote conscious experience. On that usage, the expression "conscious experience" is pleonastic and the phrase "unconscious experience" self-contradictory. But others, myself included, use the term "experience" to cover only sensory state, the kind of mental state that possesses distinctive sensory qualities. On that usage, it is not analytic that every experience is conscious. Nor is it self-contradictory that some experience is unconscious. At least, it is not obvious that either one holds. Consider, for instance, pain. It is arguable that one may not feel one's own pain if it is sufficiently slight or one is overwhelmingly distracted (LYCAN 1997; ROSENTHAL 1991). Since the objection trades on an ambiguity, it does not succeed.

The second objection is this. How could blindsight subjects have visual experiences? *They* think they do not. Why do *we* think they do?

The objection amounts to the following. Blindsight subjects think they do not have such and such visual experiences; therefore, they do not have such and such visual experiences. There is an unstated assumption here that sincere present-tense self-ascriptions of the presence or absence of visual experiences must be true. Since this infallibility thesis is highly tendentious, the first, rhetorical question could safely be put aside.

It may be thought that the objection does not have to appeal to the infallibility thesis. It does not have to assume that experiential self-ascriptions must be true. It has to assume only that such self-ascriptions must be *presumed* to be true. That is, it is necessary that, for any such self-ascription, one shall treat it as if it were true unless or until one has sufficient grounds to the contrary.

Now this thesis is much more plausible than the infallibility thesis. And it may well be true. But the presumption of truth that attaches to avowals and disavowals of experiences is consistent with our contention that blindsight subjects do in fact have the relevant visual experiences. This is on a par with the fact that the presumption of innocence is compatible with the view that many defendants are in fact guilty. The presumption of truth accorded the experiential self-ascriptions of blindsight subjects could rightly be overturned. Indeed, the first step of our argument presents behavioral evidence and theoretical considerations for doing just that. For it shows that the remarkable accuracy of blindsight subjects' avowed guesses must be underlined by a range of veridical experiences, or a range of true quasi-doxastic states that in turn must be underlined by a range of veridical experiences.

But this leads to the third objection, which is that if we overturn the presumption of truth accorded a subject's self-ascriptions of visual experiences, we should overturn the presumption of truth accorded his self-ascriptions of guesses as well. If we do not accept his claim that he does not see anything, we should not accept his claim that he only guesses. After all, self-ascriptions of guesses are not more reliable than self-ascriptions of (the absence of) visual experiences. One occupies a certain intentional state when one guesses that things are thus and so. One occupies a certain sensory state when one has a particular visual experience of things. And by and large self-ascriptions of intentional states are not more apt to be true than self-ascriptions of sensory states. But if we do not accept a blindsight subject's claim that he only guesses, we have no reason to believe in the phenomenon of blindsight. Since we have reason to believe in this phenomenon, we should not deny that a blindsight subject lacks the relevant visual experiences.

Now granted that both self-ascriptions of visual experiences and self-ascriptions of guesses enjoy a presumption of truth, there is still a difference between the two in the case of blindsight. For we have sufficient grounds for reversing the presumption of truth that attaches to a blindsight subject's self-ascriptions of visual experiences, but the same does not hold for his self-ascriptions of guesses. True, the fact that a blindsight subject made virtually no errors when he was asked to guess may give us some reason to doubt that he really guessed as he claimed to have done. But this doubt, if existent, should be dispelled once we learn that the subject expressed great surprise when he was told how well he had done, and that he was unable to explain how he had managed to do so well.

The subject was surprised because he first believed (incorrectly) that he had done badly, then came to believe just the reverse, and finally recognized the contrast between the two. Clearly, given the context, he would not have held the original belief if his responses had not been guesses. The presumption that the subject made guesses is further strengthened by the fact that he could not say how he had done so well. To be sure, performing a task well does not imply possessing an ability to explain how one does it. For the former may just be a case of exercising a practical skill whereas the latter must be a case of possessing propositional knowledge, and "knowing how" does not require "knowing that". Many people, in particular children, dance or swim well without being able to say how they do that. Such failure does not mean that they are so unsure if they are doing things right that what they are doing when they are swimming or dancing well is nothing but guesswork. But the blindsight case is different. Speaking and listening comprehension aside, what the blindsight subject did when he answered the questions about objects and events within his blind fields could not plausibly be construed as exercising a skill, practical or otherwise. And given his intellectual maturity plus the fact that what was being asked of him was not some fancy neurophysiological or mentalistic explanation, his failure to explain how he managed to answer so well in forced-choice situations very likely indicates that he did not have the relevant facts at the time and so was not reasonably sure if his answers were correct, which is just another way of saying that he only guessed. All in all, the fact that a blindsight subject made virtually no errors when he was asked to guess is explained not by the hypothesis that he did not guess at all, but by the hypothesis that he had unconscious experiences.

This brings us to the fourth objection, which is that that is still not the best explanation of the evidence. What the blindsight phenomenon shows is that information can be carried by unconscious states and processes. The nervous system is undoubtedly brimming with information received and retained, processed and transmitted, in unconscious form. But just because blindsight subjects manage to provide reliable information about objects and events within their blind fields, that does not mean that they have visual experiences. If it did, then purely mechanical devices such as sensors and detectors, meters and gauges, would be able to have experiences too—just unconscious experiences! After all, when properly designed and operated, these devices also provide reliable information about the relevant objects and events.

Suppose that an event carries information about another if and only if the first nomologically co-varies with the second in the sense that it is a law that under the normal conditions the first occurs just in case the second does (DRETSKE 1997b, 1988, 1981). Then information-carrying systems are everywhere.

Author's address

A. Minh Nguyen, Department of Philosophy, University of Louisville, 308 Humanities Building, Louisville, KY 40292, USA. Email: minh.nguyen@louisville.edu

A thermometer carries information about the ambient temperature, a tire gauge carries information about the pressure of the air in a tire, smoke carries information about fire, the number of rings in a tree carries information about its age, and so on. But if what makes blindsight subjects subjects of experience is the fact that they provide reliable information about certain things, then there is no explanation of why only mental entities are subjects of experience. For many non-mental entities perform such a function and so would count as subjects of experience by the above standard, but clearly they are not.

Now I do not hold that an entity is a subject of experience just in case the entity provides reliable information about a certain domain. Rather, my position is that an entity is a subject of experience just in case the entity has an experience. Since an experience is a sensory state, since a sensory state is a mental state, and since a mental state is a state that only mental entities occupy, only mental entities are therefore subjects of experience. So my position does not entail that purely mechanical devices and other purely physical systems can be subjects of experience. As a matter of fact, it rules out that possibility.

What justifies the claim that blindsight subjects have the relevant visual experiences is not simply that they provide reliable information about certain things, something that speedometers and tree rings can do. Rather, the justification also has to do with the fact that blindsight subjects *guess* right most of the time in a forced-choice situation, something that speedometers and tree rings cannot do because guessing is a cognitive activity and non-mental entities cannot engage in any such. This fact provides the starting point of my argument. The first step of my argument attempts to show that the remarkable accuracy of blindsight subjects' avowed guesses must be underlined by a range of veridical experiences, or a range of true quasi-doxastic states that in turn must be underlined by a range of veridical experiences. This seems to hold for creatures capable of both conceptualizing and experiencing, which blindsight subjects clearly are.

Acknowledgement

I would like to thank Frederick DRETSKE, Robert HASKELL, Avery KOLERS, Douglas MEEHAN, David ROSENTHAL, and David SMITH for their valuable comments on an earlier draft. I would also like to thank Nhi HUYNH for her constant support and encouragement.

Notes

1 A perceptual field, relative to a creature and a sensory modality, consists of all the points in space that can be perceived by that creature, in that sensory modality, from a particular position in space-time.

2 See STOERIG/COWEY (1990). Recent studies also suggest that blindsight subjects can recognize and discriminate certain types of affective facial expression within their blind fields. See HEYWOOD/KENTRIDGE (2000) and DE GELDER et al. (2000).

3 It seems universally acknowledged that a creature cannot experience or perceive a thing without being conscious; and that for a creature *C*, as opposed to a mental state, to be conscious is for *C* to be awake, as well as mentally sensitive and responsive to environmental stimuli, in the sense that *C* is not soundly asleep, whole-bodily anesthetized, deeply comatose, or similarly afflicted. "Creature consciousness" and "state consciousness" are ROSENTHAL's (2003) terminologies.

References

Block, N. (1995) On a confusion about a function of consciousness. Behavioral and Brain Sciences 18:227–247.

De Gelder, B./Vroomen, J./Pourtois, G./Weiskrantz, L. (2000) Affective blindsight: Are we blindly led by emotions? Trends in Cognitive Science 4:126–127.

Dennett, D. (1995) The path not taken. Behavioral and Brain Sciences 18:252–253.

Dretske, F. (1981) Knowledge and the flow of information. MIT Press: Cambridge MA.

Dretske, F. (1988) Explaining behavior. MIT Press: Cambridge MA.

Dretske, F. (1997a) What good is consciousness? Canadian Journal of Philosophy 27:1–15.

Dretske, F. (1997b) Naturalizing the mind. MIT Press: Cambridge MA.

Heywood, C./Kentridge, R. (2000) Affective blindsight? Trends in Cognitive Science 4:125–126.

Lycan, W. (1997) Consciousness as internal monitoring. In:

Block, N. et al. (eds) The nature of consciousness. MIT Press: Cambridge MA, pp. 755–771.
Reber, A. (1985) The penguin dictionary of psychology. Penguin: New York.
Rosenthal, D. (1991) The Independence of consciousness and sensory quality. In: Villanueva, E. (ed) Philosophical issues 1: Consciousness. Ridgeview: Atascadero CA, pp. 15–36.
Rosenthal, D. (2003) Consciousness and mind. Clarendon Press: Oxford.
Stich, S. (1978) Beliefs and subdoxastic states. Philosophy of Science 45:499–518.
Stoerig, P. (1997) Phenomenal vision and apperception: Evidence from blindsight. Mind and Language 12:224–237.
Stoerig, P./Cowey, A. (1990) Wavelength sensitivity in blindsight. Nature 342:916–918.
Weiskrantz, L. (1986) Blindsight. Clarendon Press: Oxford.

Ariane Bazan/Filip Geerardyn/Veroniek Knockaert/David Van Bunder/
Gertrudis Van de Vijver

Language as the Source of Human Unconscious Processes

Abstract

A neuropsychoanalytic framework is proposed for the study of unconsciously determined human behavior as expressed in psychic symptoms and dreams. First, some clinical observations are operationalized in an analytical FREUDO–LACANIAN perspective. In particular the notion of the human unconscious as a linguistically structured dynamic system is presented. Second, these psychoanalytical notions are integrated with current neuroscientific insights on language.
This framework essentially conceives human language as the one object of two evolutionarily radically different neurological processing circuits, acting partially in parallel. The oldest pathway processes the "objective" or phonemic qualities of language input subcortically while the second and typically human pathway processes language neocortically on its semantic qualities. The affective processing of raw phonemic material therefore is thought to operate in relative autonomy from the semantic processing and thereby able to induce so-called "false connections". It is further proposed that (1) meaningful access to language is essentially a(n articulatory) motor event, (2) imagined speech also induces this motor activation and (3) unspoken phonemes give rise to "linguistic phantoms". In final, a structural hypothesis for the FREUDO–LACANIAN unconscious is proposed conceiving this system as a raster of latent phonemic phantoms, eventually functioning as "attractors" for the subject's affective attention.

Key words

Language, unconscious, phonemes, emotion, signifier, motor imagery

Introduction

Major advances in the understanding of language dynamics have been achieved the last ten years and continue to come to us each day, originating from disciplines as diverse as medicine, neuroscience, evolutionary biology and phenomenological research in language development. The present contribution has no claim in the further elaboration of these domains, but proposes a more clinical or psychological approach to language dynamics. In clinical work, as in everyday psychopathology, the acute concern is not so much the exact neuronal trajectory of language, but far more the way a subject's emotional experience is influenced by or has influence on his or her particular language dynamics. The focus is therefore precisely this emotional language processing and the way it is proposed to be at the origin of the typically human unconscious mind.

Problem Presentation

Clinical observations

There has been, since FREUD (1960, 1975) and in particular with LACAN (1957), a particular interest and attention paid to the literal language patients use when talking about themselves or their problems in consultation. This is illustrated in FREUD's clinical oeuvre from the start with, for example, his illustrative cases presented in the "Psychopathology of Everyday Life" (FREUD 1960). The forgetting of the name "Signorelli" for example seemed not to be motivated by some conflicting semantics connected with the painter "Signorelli", but, curiously, with the semantics of a phonological variant of the word "Signorelli", in this case *signor* or "master". In "The interpretation of dreams" FREUD (1975) introduced the concept that dreams are frequently to be taken literally and that these literal transcripts are subsequently to be read as rebuses. That the same principle also pertains to symptoms, as FREUD proposes, is clearly illustrated in a letter to FLIESS (FREUD 1986, pp316–331) which is given here as a paradigmatic example:

"A little interpretation came my way… Mr. E. had an anxiety attack at the age of ten when he tried to catch a black beetle… The meaning of this attack had thus far remained obscure. Now, dwelling on the theme of "being unable to make up one's mind", he repeated a conversation between his grandmother and his aunt about the marriage of his mother… from which it emerged that she had not been able to make up her mind for quite some time; then he suddenly came up with the black beetle, which he had not mentioned for months, and from that to ladybug [Marienkäfer] (his mother's name was Marie); then he laughed out loud… Then we broke off and next time he told me that before the session the meaning of the beetle [Käfer] had occurred to him; namely: que faire? = being unable to make up one's mind… meschugge!

"You may know that here a woman may be referred to as a nice "beetle". His nurse and first love was a French woman; in fact, he learned to speak French before he learned to speak German".

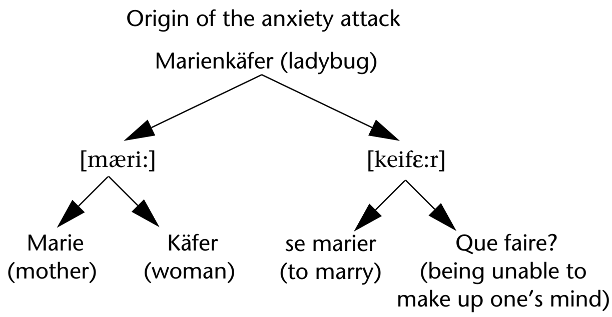

Figure 1. "The inability of mother to make up her mind concerning her marriage" = origin of the (current) anxiety.

As is made clear in the scheme on Figure 1, it seems that the literal forms of the words function as carrier of affects, more or less independently of their semantics.

In the neuropsychoanalytic research unit at the University of Ghent a number of clinical observations of this kind were systematically recorded in collaboration with diverse clinicians. Three typical examples of this phenomenon are presented briefly: (1) a dream: a woman dreams that she is sitting in front of her therapist and that their feet are touching; the meaning of the dream becomes clear when she formulates its content as "we sat sole to sole"; (2) an anxiety: a woman gets an anxiety attack when her friend, promising her a hot time together, whispers to her: "Je te montrerai les sommets de la merveille" ("I will show you the top of the record"); upon analysis, it seems that the phonological carrier [læ mɛrvɛ:j] was for the young woman also referring to "la mère veille", i.e., mother is watching; (3) a dream: a pregnant woman dreams she is driving a big Mercedes down a spiral driveway; while she is driving the car, the driveway gets narrower and at one point her car gets stuck; upon analysis, it seems that the phonological carrier [læ mɛ:rsɛ:dɛs] was for the woman also referring to: "la mère cède" (i.e., "the mother fails"). The woman was at that moment preparing for the presentation of her Ph.D. thesis and therefore experiencing some conflict between this energy consuming achievement and her imminent motherhood. Constant in these examples is that the origin of the symptom or dream is phonologically—and not semantically—related to the actual form in which the dream or symptom presents itself.

Psychoanalytical framework

In a FREUDO–LACANIAN framework the reference to a human unconscious refers to the idea of an unconscious that is structured like a language (LACAN 1972–1973; VAN BUNDER et al. 2002), and to the dynamic system thought to be at the origin of signifier mediated affective "mismatches" as illustrated in the aforementioned examples. For FREUD (1978), the word form or "word-representation" implicates an acoustic component, "the acoustic image" and a motor component or "speech movement representation", the kinesthetic incoming information of the articulatory system. This word-representation level has therefore a finite number of components and is as such to be distinguished from the "object-representation" level. This object-representation level has an infinite number of components, including e.g., visual, acoustic and tactile recordings of the object. The similarity of this model with the model of DAMASIO et al. (1996) and CARAMAZZA (1996) is remarkable. The word-representation level can be considered as corresponding to the lexical level, to be situated in the left basal temporal lobe, while the object-representation level obviously corresponds with the semantic level, to be situated distributed over the temporo-parieto-occipital areas of both hemispheres. For LACAN (1957), the word form or "signifier" is a phonemic carrier in the SAUSSURIAN sense, without any predetermined signification. The signifier is an "empty" lexical position depending on the context for its conscious signifying. In his later seminars (LACAN 1972–1973; see also VAN DE VIJVER 2002), the signifier is not any longer this empty or neutral but is carrier of an unconscious sense through its phonemic articulation with the body.

Two Evolutionary Pathways for Language

The idea that language is also a carrier of sense *apart from its semantic sense* is implicit in these clinical examples above: it seems to be carrier of affect, in particular of anxiety or arousal, independently of its semantics, i.e., the phonemic carrier induces bodily changes on an affective level, apart from the access to semantics. Language therefore seems to be the carrier of two more or less independent levels of signification: a semantic one and an affective one.

The possibility of relative independence between the affective and "scenic" (or "declarative") content of the same input material is actually central to LEDOUX' theory on emotional processing (LEDOUX 1993, 1994). Central to his view is the wedge-like splitting of the neuronal trajectory of a single input train into two categorically different pathways, one subcortical or limbic and the other at the level of the neocortex. The limbic trajectory is both phylogenetically old and ontogenetically early: the systems are functional from birth on (and probably earlier) and immediately start establishing an emotional memory on the basis of conditioning of raw input material (LEDOUX 1993, 1994). The neocortical trajectory is both phylogenetically more recent and ontogenetically late: cortical maturation is not achieved until six to ten years after birth. Therefore, it is only with some delay that an articulate mature "cognitive" analysis of the input material can be fully achieved and stored in the semantic fields.

LEDOUX' scheme has some remarkable similarities with FREUD's idea of the "splitting of consciousness" as he formulates it in "The neuro-psychoses of defense" (FREUD 1961, p51–52): "If someone with a disposition [to neurosis] lacks the aptitude for conversion, but if, nevertheless, in order to fend off an incompatible idea, *he sets about separating it from its affect*, then that affect is obliged to remain in the psychical sphere. *The idea*, now weakened, is still left in consciousness, separated from all association. But *its affect*, which has become free, attaches itself to other ideas which are not in themselves incompatible; and thanks to this "false connection", those ideas turn into obsessional ideas".

The notion of the "splitting of consciousness" thus implies the splitting of an idea or experience in its content on the one hand and its affect or excitation sum on the other. The sum of excitation is invested in body innervations in conversion hysteria, or into other ideas in obsessional neurosis. The principle, however, remains the same: one and the same experience can psychologically be conceived as a "complex" of separable elements, with different dynamic characteristics, different fates and different output systems, which, without being completely independent from each other, nevertheless possess a relative autonomy.

It is therefore tempting to explain the above illustrated signifier mediated mismatches in a similar manner. Language is as appropriate an input stimulus as another (a non-language auditory or a visual stimulus) and is therefore also considered to be subject to "emotional conditioning". This emotional conditioning is relatively independent from semantics, which is considered an operation of the higher associative neocortical areas. At the level at which language is thought to be emotionally conditioned, clinical work teaches us that there is no difference between "soul" and "sole". Language is at that level not treated as a fundamentally ambiguous system that has to be contextually interpreted, but as any other object, i.e., unambiguously or objectively. Like other objects, the language object automatically activates a number of proper sensory and motor associations correlated with its particular phonemic form (and not with its semantic meaning) and the emotional activation is thought to be effective at the level of these, most probably, motor associations.

Incoming Phonemes Are Motor Programs

"Affective mismatch" operates at the phoneme level

Clinical work teaches us more than that. First, it seems clear that the substratum for emotional activation is not the raw acoustic material of language, but obviously its phonemic transcription. What is effective in eliciting an emotional activation does not seem to be necessarily endowed with some particular acoustic qualities; apparently, what does seems important, however, are the phonemic invariants of the message. Second, it appears that this emotional processing mechanism does not seem to respect word boundaries. In "*la merveille*", e.g., the relevant activating substratum ([læ merve:j]) can either be packed in one word or in a complete sentence.

These indications help us to speculate on the exact nature of the physiological language substratum responsible for the emotional activation. In a comprehensive model for the neural basis of auditory sentence processing FRIEDERICI (2002) situates the process of identification of phonemes as the second step, immediately after the primary acoustic analysis. In her scheme, this identification of phonemes involves a projection from the left temporal BRODMANN area 42 (adjacent to HESCHL's gyrus or the primary auditory cortex) to the left prefrontal BRODMANN area 44 or BROCA area and is completed within the first 100 ms of auditory processing. Immediately after this step, a number of lexical operations are then exerted on the phonemic material, implying between others the identification of word form and of word category. As neither word form, nor word category, nor, *a fortiori*, any semantic identification, is relevant in the described signifier-mediated affective mismatches, we speculate that emotional processing responsible for these mismatches takes place after phonemic processing but before lexical processing of language, i.e., during the first 100 ms after presentation.

Phoneme identification involves motor activation

Even if in the literature there might be some confusion about the exact extent of implication and about the exact brain locus of interest (for a review, see BURTON 2001), there seems to be large agreement that phonemic identification does involve motor areas situated either prefrontally in the BROCA area (HICKOK/POEPPEL 2000) or subcortically, i.e., implicating basal ganglia and/or cerebellar pathways (IVRY/JUSTUS 2001). This observation therefore gives weight to LIBERMAN's motor theory of speech perception" (LIBERMAN et al. 1967; LIBERMAN/MATTINGLY 1985). This theory, based on phonological research, holds that the basis of speech perception is not the actual sound of speech, but rather the "articulatory gestures" made by the speaker. It argues that listeners identify spoken words through using that information to access their speech motor system. This is supported by the fact that the speech phones, the smallest units we can hear in words, link to articulatory and not auditory-related invariants. Phoneticians classify and characterize phones nearly entirely in terms of how they are articulated and not in terms of how they sound. The theory thereby accounts for our ability to perceive the invariant articulatory events that form the speech stream, in spite of the great variability in the acoustic signal.

There has been a more recent neural instantiation of this motor theory by RIZZOLATTI/ARBIB (1998). These researchers report that in monkeys a part of the premotor cortex (F5) contains neurons that discharge both when the monkey grasps or manipulates objects and when it observes the experimenter making similar actions. Recent studies suggest that a similar system exists in humans. FADIGA et al. (1995) reported evidence for motor activation when human subjects merely observed an action, and the muscles activated were those that would have been used had they performed the action themselves. RIZZOLATTI/ARBIB (1998) also show that there are neurons in F5 in the monkey's brain that respond both when the animal makes lipsmacking movements and when it observes them in others. Of particular importance is the fact that these authors note that area F5 in the monkey is the probable homologue of BROCA's area in humans. ZATORRE et al. (1992, 1996) have indeed argued that the mapping of the incoming speech stream onto the linguistically relevant units, which are thought to be the corresponding articulatory gestures, activates BROCA's area. There is some parallel argumentation, especially coming from CORBALLIS (1999), that the origins of human language might be situated in manual gesture rather than in vocalization. Recently, CALLAN et al. (2002) have shown that the presence of such mirror neurons in speech motor areas of the brain may explain why lip-reading enhances the intelligibility of what a person is saying. This finding adds strength to the argument that human speech evolved from a primitive gestural system of communication, rather than from simple vocalizations. For all these reasons, RIZZOLATTI/ARBIB (1998) propose that the development of the human speech circuit is a consequence of the fact that the precursor of BROCA's area was endowed, before speech appearance, with a mechanism for recognizing actions made by others.

This idea of perception–action linking already stood central in Freud's "Project for a scientific psychology" (FREUD 1995, pp333–334): "While one is perceiving the perception, one copies the movement oneself—that is, one innervates so strongly the motor image of one's own which is aroused towards coinciding [with the perception], that the movement is carried out. Hence one can speak of a perception having an *imitation-value*. (…) Thus judging, which is later a means for the *cognition* of an object that may possibly be of practical importance, is originally an associative process between cathexes coming from outside and arising from one's own body—an *identification of information or cathexes from Θ [the perception] and from within*". More generally, this suggests that external stimulation only makes sense for the brain if reprocessed into something self-initiated (GEERARDYN 2002). In his study on aphasia, FREUD (1978, pp91–92) then suggested that in language this movement might be thought of as articulation: "Understanding of spoken words is probably not to be regarded as simple transmission from the acoustic elements to the object association; it rather seems that in listening to speech for understanding, the function of verbal association is stimulated from the acoustic elements at the same time, so that *we more or less repeat ourselves the words heard*, thus supporting *our understanding with the help of kinaesthetic impressions*. A higher measure of attention in listening will entail a higher degree of transmission of speech heard on to the tract serving the motor execution of language".

A comprehensive framework for the mechanism of "affective mismatches"

In summary, I suggest that in the signifier-mediated affective mismatches as illustrated above, the phonemic transformation of the incoming linguistic material is the effective substratum and that this phonemic transformation involves language motor pathways, and therefore that the significant "affective mismatch" is to be situated at the motor–limbic interface. The full mechanism of these affective mismatches is understood as follows. Any time affectively colored phonemes are actualized in the ongoing discourse, be it not in the right original semantic context, affect is nevertheless aroused and may be falsely connected to the actual semantics (e.g., the anxiety aroused by the beetle is falsely attributed to the appearance of the beetle). The selection process for the pertinent semantic interpretation can be conceived as an active inhibitory process, which "represses" contextually non-valid semantic alternatives (cf. SIMPSON/KANG 1994; FAUST/GERNSBACHER 1996; GORFEIN/BERGER/BUBKA 2000). Since, however, affective activation is thought not to be subject to this cortical inhibition process (cf. the automaticity of affect, DE HOUWER/EELEN 1998; FAZIO 2001), it may be the case that this irrepressible affect is experienced in the "wrong" semantic context and therefore gives rise to falsely connected symptoms in psychopathology.

Hypothesis: The Unconscious is Affect Aroused by Phonemic "Phantoms"

Imagined speech is motor activation.

Speech motor areas are not only activated in case of speech production or active speech perception, but in a number of other conditions where speech is imagined but not effectively produced, including inner speech (MCGUIRE et al. 1996), auditory verbal imagery (MCGUIRE et al. 1996) and hallucinations in schizophrenia (LIDDLE et al. 1992; CLEGHORN et al. 1992; MCGUIRE/SHAH/MURRAY 1993). As imagining movements leads to increased cerebral blood flow in motor areas concerned with their execution (DECETY et al. 1994; STEPHAN et al. 1995), the activity in regions which control speech motor systems may be due to imagined speech in these conditions.

DECETY/GRÈZES (1999) define motor imagery as a dynamic state during which the representation of a given motor act is internally rehearsed within working memory without any overt motor output. It has been proposed that such a simulation process corresponds to the conscious counterpart of many situations experienced in everyday life, such as watching somebody's action with the desire to imitate it, anticipating the effects of an action, preparing or intending to move, refraining from moving, and remembering an action (JEANNEROD/DECETY 1995; DECETY 1996).

All of these tasks involve motor representations that recruit neural mechanisms specific to action planning. GEORGIEFF/JEANNEROD (1998) and DECETY/GRÈZES (1999) remark that comparison of brain activation during several modalities of action representation (including observation and imagining) reveals a common network to which the inferior parietal lobule (area 40), part of the supplementary motor area (SMA), the ventral premotor area, the cingulate gyrus and the cerebellum contribute. The ventral premotor area corresponds to a crossroads between the ventral part of area 6 and areas 44 and 45 (BROCA's area), a cortical zone which bears some homology with the monkey ventral area 6 where mirror neurons are recorded (RIZZOLATTI et al. 1996). For all these reasons, it is expected that, similar to what is proposed for the signifier mediated affective mismatches, the substratum for speech imagery also involves the phonemic motor pathways.

Imagined motor activation induces phantoms

The phantom limb syndrome refers to the strong perception that a missing limb is still there, to the sense of being able to move it and to reported feelings arising from it such as intense pain. RAMACHANDRAN (1994) suggests that the relevant signaling maintaining the phantom are the sensations arising from reafference signals derived from the motor commands sent to the muscles of the phantom. Pain could especially be linked to the missing of the corresponding sensory feedback that would confirm the movement execution. The concepts of "efference copy", as introduced by HELMHOLTZ (1995) suggests that a copy of one's *intended* movement is used every time a voluntary action is planned, such that the sensory consequences of the action can be anticipated and eventually cancelled (BLAKEMORE et al. 1998). Most contemporary accounts of efference copy have claimed that it is unconscious, or acts to cancel percepts rather than generate them. Nevertheless, some clinical and experimental observations suggest that this information, in particular the state of the motor system, can influence subjective perception of the body. In deafferentiation, people still gesticulate while talking, even when the sight of these gestures is blinded to them and their interlocutor (COLE/PAILLARD 1996). It is suggested that, in agreement with MERLEAU-PONTY (1945) and IVERSON/GOLDIN-MEADOW (1998), the gesticulation when talking is for the subject's own linguistic-thought processes and not just for communicative purposes and that the informative signals here are not the (absent) somato-sensory signals but, remarkably, the efference copies of the hand muscles. MCGONIGLE et al. (2002) recently report the case of E. P., a right-handed female stroke patient with a right frontomesial lesion who sporadically experiences a supernumerary "ghost" left arm. Their results suggest that areas traditionally classified as part of the motor system can influence the conscious perception of the body and they propose that, as a consequence of her injury, E. P. is aware of the position of the phantom limb in its "action space" on the basis of the efference (motor) copies while also continuing to be aware of the true position of her real limb on the basis of afferent somato-sensory information. Focusing on the desired goal of an action, JEANNEROD (1994, p201) suggests that neurons encoding the "final configuration" of the body would continue firing "until the goal has been reached". If the goal were not reached, "the sustained discharge would be interpreted centrally as a pure representational activity and give rise to mental imagery" (JEANNEROD 1994, p201). In the case of E.P. sustained activity in a traditionally motor area of the brain (the SMA) correlates with her perception of a phantom arm (MCGONIGLE et al. 2002). All these observations suggest that phantoms arise when motor commands are consistently given, and that phantom pain could especially be linked to the missing of confirmatory sensory feedback.

Hypothesis: A linguistic unconscious

As a hypothesis, it is therefore proposed that recurrent unspoken phonemes, the motor circuitry of which is regularly activated, either by directed speech perception, by linguistic imagery or by refraining from speaking them ("repression"), could similarly create "phantom" phonemes in a linguistic "action space". These "phantom phonemes" which would be often or lastingly internally rehearsed in the phonological loop of working memory, would, if not spoken (enough), be interpreted centrally as representational activity, giving rise to mental imagery. This mental imagery then is not to be conceived as primarily semantically structured, but would rather have a primarily phonemic structure.

It is difficult to further speculate on the nature of such a phonemic mental imagery, but as phonemes were proposed to be affect-carrying substrates, it makes sense to conceive that although executing the motor plan is subject to inhibition, the associated affective activation is not. In this perspective, it is interesting to note that motor imagery activates heart and respiration control mechanisms in proportion to the actual effort that would be required for the real action (DECETY et al. 1991; DECETY et al. 1993; WANG/MORGAN 1992). Such an autonomic response in a situation where no muscular work is produced can only be attributed to a central influence similar to that observed during motor preparation. As the autonomic system is by definition independent of voluntary control and cannot be held under inhibition, central influences on this system become recordable at the periphery (JEANNEROD 1994).

Author's address

Ariane Bazan, Psychiatric Center, Sint-Amandus, Beernem, Belgium, and Research Unit in Neuro-Psychoanalysis at the Department for Psychoanalysis and Clinical Consulting, University of Ghent, Belgium.
Email: ariane.bazan@rug.ac.be

In conclusion, we propose the following structural hypothesis regarding the FREUDO–LACANIAN unconscious: phonemes of particular importance in one's personal history (e.g., the proper name) need not to be actualized per se in the ongoing discourse but have formed throughout the individual's particular history a raster of phonemic "phantoms", that are continuously functioning as "attractors" for the subject's affective attention.

References

Blakemore, S. J./Wolpert, D. M./Frith, C. D. (1998) Central cancellation of self-produced tickle sensation. Nature Neurosciences 1:635–640.

Burton, M. W. (2001) The role of inferior frontal cortex in phonological processing. Cognitive Science 25:695–709.

Callan, D./Jones, J./Munhall, K./Kroos, C./Callan, A./Vatikiotis-Bateson E. (2002) Mirror neuron system activity and audiovisual speech perception. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-Rom in NeuroImage 16(2).

Caramazza, A. (1996) The brain's dictionary. Nature 380:485–486.

Cleghorn, J. M./Franco, S./Szetchtman, B./Kaplan, R. D./Szetchtman, H./Brown, G. M./Nahmias, C./Garnett, E. S. (1992) Towards a brain map of auditory hallucinations. American Journal of Psychiatry 149:1062–1069.

Cole, J. D./Paillard, J. (1996) Living without touch and peripheral information about body position and movement: Studies upon deafferented subjects. In: Bermudez, A. M./Eilan, N. (eds.) The body and the self. MIT/Bradford Press: Cambridge, MA, pp. 245–266.

Corballis, M. C. (1999) The gestural origins of language. American Scientist 87:138–145.

Damasio, H./Grabowski, T. J./Tranel, D./Hichwa, R. D./Damasio, A. R. (1996) A neural basis for lexical retrieval. Nature 380:499–505.

Decety, J. (1996) Neural representations for action. Reviews in Neurosciences 7:285–297.

Decety, J./Grèzes, J. (1999) Neural mechanisms subserving the perception of human actions. Trends in Cognitive Sciences 5:172–178.

Decety, J./Jeannerod, M./Durozard, D./Baverel, G. (1993) Central activation of autonomic effectors during mental simulation of motor actions in man. Journal of Physiology 461:549–563.

Decety, J./Jeannerod, M./Germain, M./Pastene, J. (1991) Vegetative response during imagined movement is proportional to mental effort. Behavioral Brain Research 42:1–5.

Decety, J./Perani, D./Jeannerod, M./Bettinardi, V./Tadary, B./Woods, R./Mazziotta, J. C./ Fazio, F. (1994) Mapping motor representations with positron emission tomography. Nature 371:600–602.

De Houwer, J./Eelen, P. (1998) An affective variant of the Simon paradigm. Cognition and Emotion 12:45–61.

Fadiga, L./Fogassi, L./Pavesi, G./Rizzolatti, G. (1995) Motor facilitation during action observation: A magnetic stimulation study. Journal of Neurophysiology 73:2608–2611.

Faust, M. E./Gernsbacher, M. A. (1996) Cerebral mechanisms for suppression of inappropriate information during sentence comprehension. Brain and Language 53:234–259.

Fazio, R. H. (2001) On the automatic activiation of associated evaluations: An overview. Cognition and Emotion 15:115–141.

Friederici, A. D. (2002) Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences 6:78–84.

Freud, S. (1960) The psychopathology of everyday life. The standard edition of the complete psychological works of Sigmund Freud (Translated and edited by J. Strachey), Volume 6. The Hogarth Press: London. Originally published in 1901.

Freud, S. (1961) The neuro-psychoses of defence. In: Strachey, J. et al. (eds) The standard edition of the complete psychological works of Sigmund Freud, Volume 3. Hogart Press: London, pp. 62–70. Originally published in 1894.

Freud, S. (1975) The interpretation of dreams. The standard edition of the complete psychological works of Sigmund Freud, volumes 4–5 (Translated and edited by J. Strachey). The Hogarth Press: London. Originally published in 1900–1901.

Freud, S. (1978) On aphasia: A critical study (Translated by E. Stengel). International Universities Press: New York. Originally published in 1891, pp. 91–92.

Freud, S. (1986) Briefe an Wilhelm Fließ 1887–1904 (Ungekürzte Ausgabe). Fischer Verlag: Frankfurt am Main. Originally published in 1897.

Freud, S. (1995) Project for a scientific psychology. In: Strachey, J. et al. (eds) The standard edition of the complete psychological works of Sigmund Freud, Volume 1. The Hogart Press: London, pp. 281–397. Originally published in 1895.

Geerardyn, F./Van de Vijver, G./Knockaert, V./Bazan, A./Van Bunder, D. (2002) "How do I know what I think till I hear what I say": On the emergence of consciousness between the biological and the social. This issue.

Georgieff, N./Jeannerod, M. (1998) Beyond consciousness of external reality. A "Who" sytem for consciousness of action and self-consciousness. Conscious Cognition 7(3):465–77.

Gorfein, D. S./Berger, S./Bubka, A. (2000) The selection of homograph meaning: Word association when context changes. Memory and Cognition 28:766–773.

Helmholtz, H. (1995) Science and culture: Popular and philosophical essays (Edited and translated by D. Chahan). University of Chicago Press: Chicago.

Hickok, G./Poeppel, D. (2000) Towards a functional neuroanatomy of speech perception. Trends in Cognitive Sciences 4:131–138.

Iverson, J. M./Goldin-Meadow, S. (1998) Why people gesture when they speak. Nature 396:228.

Ivry, R. B./Justus, T. C. (2001) A neural instantiation of the motor theory of speech perception. Trends in Neurosciences 24:513–515.

Jeannerod, M. (1994) The representing brain: Neural correlates of motor intention and imagery. Behavioral and Brain Sciences 17:187–245.

Jeannerod, M./Decety, J. (1995) Mental motor imagery: A window into the representational stages of action. Current Opinion in Neurobiology 5:727–732.

Lacan, J. (1975) Le séminaire. Livre XX: Encore (1972–1973). Seuil: Paris.

Lacan, J. (1999) L'instance de la lettre dans l'inconscient ou la raison depuis Freud. In: Lacan, J., Ecrits I (2ième ed. de poche). Seuil: Paris, pp. 490–526. Originally published in 1957.

LeDoux, J. E. (1993) Emotional memory systems in the brain. Behavioural Brain Research 58:69–79.

LeDoux, J. E. (1994) Emotion, memory and the brain. Scientific American 6:32–39.

Liberman, A. M./Cooper, F. S./Shankweiler, D. P./Studdert-Kennedy, M. (1967) Perception of the speech code. Psychological Review 74:431–461.

Liberman, A. M./Mattingly, I. G. (1985) The motor theory of speech perception revised. Cognition 21:1–36.

Liddle, P. F./Friston, K. J./Frith, C. D./Hirsch, S. R./Jones, T./ Frackowiak, R. S. J. (1992) Patterns of cerebral blood flow in schizophrenia. Britisch Journal of Psychiatry 160:179–186.

McGonigle, D. J./Hänninen, R./Salenius, S./Hari, R./Frackowiak, R. S. J./Frith, C. D. (2002) Whose arm is it anyway? An fMRI case study of supernumerary phantom limb. Brain 125:1265–1274.

McGuire, P. K./Shah, G. M. S./Murray, R. M. (1993) Increased blood flow in Broca's area during auditory hallucinations in schizophrenia. Lancet 342:703–706.

McGuire, P. K./Silbersweig, D. A./Murray, R. M./David, A. S./ Frackowiak, R. S. J./Frith, C. D. (1996) Functional anatomy of inner speech and auditory verbal imagery. Psychological Medicine 26:29–38.

Merleau-Ponty, M. (1945) Phenomenologie de la perception. Gallimard, Paris.

Ramachandran, V. S. (1994) Phantom limbs, neglect syndromes, repressed memories, and Freudian psychology. International Review of Neurobiology 37:291–333.

Rizzolatti, G./Arbib, M. A. (1998) Language within our grasp. Trends in Neuroscience 21:188–194.

Rizzolatti, G./Fadiga, L./Gallese, V./Fogassi, L. (1996) Premotor cortex and the recognition of motor actions. Cognitive Brain Research 3:131–141.

Simpson, G. B./Kang, H. (1994) Inhibitory processes in the recognition of homograph meanings. In: Dagenbach, D./ Carr, T. H. (eds) Inhibitory processes in attention, memory, and language. Academic Press: San Diego CA, pp359–381.

Stephan, K. M./Fink, G. R./Passingham, R. E./Frith, C. D./ Frackowiak, R. S. J. (1995) Functional anatomy of the mental representation of hand movements in healthy subjects. Journal of Neurophysiology 73:373–386.

Van Bunder, D./Knockaert, V./Van de Vijver, G./Geerardyn, F./Bazan, A. (2002) The return of the repressed. Anticipation and the logic of the signifier. International Journal of Computing Anticipatory Systems. Forthcoming.

Van de Vijver, G./Van Bunder, D./Knockaert, D./Bazan, A./ Geerardyn, F. (2002) The role of closure in a dynamic structuralist viewpoint of psychic systems. This issue.

Wang, Y./Morgan, W. P. (1992) The effects of imagery perspectives on the physiological responses to imagined exercise. Behavioural Brain Research 52:167–174.

Zatorre, R./Evans, A./Meyer, E./Gjedde, A. (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846–849.

Zatorre, R./Meyer, E./Gjedde, A./Evans, A. (1996) PET studies of phonetic processing of speech: review, replication and reanalysis. Cerebral Cortex 6:21–30.

Wilma Bucci

The Language of Emotions

An Evolutionary Perspective

Abstract

Humans must negotiate their experiential worlds as multi-system processors with substantial but limited integration of systems. Emotions are a crucial part of the information processing system for humans as for all organisms. The diversity of systems and the dissociation among them are related in interesting ways to the dimension of what has been termed unconscious processing. Systems whose formats are very different from the symbolic verbal format are likely to be experienced as unconscious or out of awareness. The phenomena that have been characterized as unconscious may be more accurately characterized as dissociated—processed in domains that we are not accustomed to accessing. The inherent limitations of psychic integration apply to varying degrees and in different forms for all people; they are not restricted to particular diagnostic categories or developmental stages. I will first provide some background for these claims, then present my theories of multiple coding and the referential process—the means by which the diverse systems may be integrated to some degree.

Key words

Multiple codes, referential process, emotions, language.

HUMANS, LIKE OTHER species, carry out complex information processing in multiple sensory, somatic and motoric modalities throughout life. Like other species, we are constantly processing the world around us in visual, auditory and other sensory modalities, and responding in multiple channels of action and somatic change. We enter a busy highway, respond to a baby's smile, recognize conditions of relative safety or danger, and generally make our way through our worlds by processing meanings and determining responses directly in perceptual, somatic and motoric modes.

Humans differ from other species, however, in that the constant flow of nonverbal processing in multiple channels is also connected to the powerful system of language to some degree. The process of expressing nonverbal experience in language, finding words for things, is an important and necessary linguistic function, that we call on continuously in everyday life. We teach our children: "don't scream, use words". The process often operates smoothly; we become aware of it primarily when it breaks down—when we have difficulty "saying what we mean". We are "struck dumb" with awe, or with horror; something is "too wonderful for words". Our popular love songs are full of expressions of this failure to connect emotion to language: "I can't begin to tell you how much you mean to me". In addition to the difficulties of communicating deep emotional experience, we become aware of the limitations of language when we try to verbalize experiences whose nuances go beyond words; when we try to describe someone's facial expression or the taste of a wine; to tell someone how to execute a skiing maneuever or dance rhumba or tango. We perceive the expression or the taste and respond to it within various internal experiential modalities; but often cannot verbalize either the perception or one's own experiential response. We can demonstrate the perceptual and motoric functions required to carry out skilled actions, but can describe them verbally to only a limited extent.

In such contexts we are forced to recognize that the connection of nonverbal experience to language is partial and limited, far more partial and limited than we may assume, or than cognitive psychologists or linguists or even psychoanalysts generally recognize. Language can express only a small proportion of the complex flow of inner experience, but language is so powerful in many respects as to mask or hide the far more pervasive nonverbal functions.

The challenge for adaptive functioning is the integration of diverse systems; humans may still be transitional in this respect. We can trace this integrative process, which I call the referential process, as occurring in parallel form on three levels: in the evolution of the human species, when the system of language with its unique design features is overlaid on processing systems with different organizing principles operating in their own formats; in development, when the same transition occurs in each child's life; and in every communicative situation, when the speaker or writer attempts to contain and communicate the operations of multiple functions with diverse formats in the single channel of the verbal code.

Diverse Views on Multiple Systems

There is considerable evidence in many fields, including evolutionary, developmental and cognitive psychology, and neuroscience, for the mind as a multi-channel, multi-system processor. All species—from bees to primates—have organized perceptual fields, organized fields of cognitive understanding, and organized modes of communication, operating outside language; this includes humans from the beginning of life, well before language has been acquired. The field of evolutionary psychology has provided us with extensive evidence concerning multiple spheres of experience in all species where such organization has been investigated (GALLISTEL 1990; KEIL 1989), and the fields of cognitive science and related disciplines have provided a wide range of models for diverse types of processing systems.

Separate continuous and categorical processing subsystems, operating in ways specific to different sensory and motoric modalities, have been found in all species. Thus we may see complex continuous analog signals that represent intensity or magnitude, as in the dance of the bee, whose liveliness varies with the richness of the food source that he is reporting to his hivemates; and categorical formats based on finite sets of calls with specific instinctual meanings, including warnings, territorial claims, or sexual interest. Wherever such bipartite functions have been examined: "in bee, pigeon, monkey, and chimpanzee, as in man—they have been found" (BORNSTEIN 1979, p54).

Within cognitive science, a number of major functional classifications have been developed, to account for the different ways that humans and other species learn, remember and think. Some of the major distinctions are:

1. Behavioral (habits and skills) vs. representational knowledge, reflected primarily in the distinction between declarative and non-declarative (or procedural) memory (COHEN 1984; SQUIRE 1992);
2. Conscious vs. unconscious knowledge as in SCHACTER's (1989) distinction between explicit and implicit memory;
3. Intentional vs. automatic processes as noted by POSNER/SNYDER (1975);
4. General knowledge vs. specific memories, as in TULVING's (1972) semantic vs episodic distinction.

These classifications overlap and intersect to varying degrees, but do not entirely correspond, as I have discussed elsewhere (BUCCI 1997). Several tripartite schemes have also been identified that incorporate components of these systems; for example TULVING (1985) added a system of "procedural memory" to his episodic–semantic distinction. Researchers have also identified modes of information processing that are specific to different sensory modalities—operating differently, for example, in processing of olfactory, visual, and auditory information.

From other perspectives, formulations concerning multiple memory systems have also been proposed, within the general framework of modularity of function, by many investigators, including MINSKY (1975), FODOR (1983), GARDNER (1983), GAZZANIGA (1985, 1988), FARAH (1984), FARAH/MCCLELLAND (1991), KOSSLYN (1987), and PINKER (1994). PINKER characterizes the human mind as incorporating multiple innate modules, or instincts, which evolved to meet different environmental demands. His definition of such instincts rests on both behavioral and physiological criteria; functions characterized as instinctual are universal within a species, and readily acquired without explicit training; they also correspond, in principle, to physiologically cohesive circuits or subsystems in the underlying brain tissue. Language is a paradigmatic example of such an instinctual module, emerging only in the human species. PINKER applies the notion of innate adapted computational modules to identify a wide range of functions, in addition to language, that might eventually be found to meet similar specifications. Possible modules of the human mind, as proposed by PINKER (1994), include the following:

1. Intuitive mechanics: knowledge of the motions, forces and deformations that objects undergo.
2. Intuitive biology: understanding of how plants and animals work.
3. Number.

4. Mental maps for large territories.
5. Habitat selection: seeking of safe, information-rich, productive environments …
6. Danger, including the emotions of fear and caution, phobias for stimuli such as heights, confinement, risky social encounters, and venomous and predatory animals, and a motive to learn the circumstances in which each is harmless.
7. Food: what is good to eat.
8. Contamination, including the emotion of disgust,… and intuitions about contagion and disease.
9. Monitoring of current well-being ….
10. Intuitive psychology; predicting other people's behavior from their beliefs and desires.
11. A mental Rolodex: a database of individuals [with blanks for features].
12. Self-concept: [data about one's value]
13. Justice: sense of rights, obligations and deserts…
14. Kinship, including nepotism and… parenting…
15. Mating, including feelings of sexual attraction, love, and intentions of fidelity and desertion.

The premise of specific innate modules and certainly the basis for their classification is even more controversial, and less well supported, for specific cognitive functions than for language; it is not my intent to examine this controversy. The point I would like to emphasize here is that the observations of PINKER and others in this field support and extend the recognition of multiplicity of function in the human system, with its evolutionary roots, while raising further questions as to how to map the mental terrain.

The theory of psychoanalysis was built on the premise of dual (or multiple) systems of thought; psychoanalysis was perhaps the first multi-modal system of mental functioning; although the characterization of the systems has been repeatedly revised. In developing his successive models of the psychical apparatus, FREUD shifted from qualities (conscious, preconscious, unconscious) to structures (id, ego, superego). (See FREUD 1964 for a summary of these theoretical developments.) He also defined the polarities of the primary vs. secondary processes of thought, related to both these dimensions. There is considerable disorder in the theory, as we know, and the theoretical correspondence among the dimensions is a matter of controversy, but the premise of multiple systems remains. We may also note the duality of the drive theory—eros and the death instincts—existing within the unconscious id zone; this is essentially orthogonal to the topographical and structural dimensions.

The Problem of Integration in the Heterogeneous Mind

The basic question must then arise; if this widely shared multi-system view of the human mind is valid; if there are multiple memory systems, as many cognitive scientists have recently shown, operating within and outside of awareness; if there are multiple formats or structures of thought as formulated a century ago within psychoanalytic theory; if there are different processing modes in the different sensory modalities; if the mind is composed of distinct modules "each keyed to the peculiar logic and laws of one domain" as PINKER has asserted (1994: 410); then how does one account for the integration of these systems in the pursuit of the individual's overall goals; how do these modules interconnect—because interconnect they must. Adaptive functioning depends on the adequate interconnection of the diverse systems or modules that populate the heterogeneous human mind. Using PINKER's elaborate systems of modules, for example, if one's goal is to select and acquire and set up a place to live one needs to call on a number of distinct functions, such as mechanics and number and mental maps and habitat selection; if one seeks to find a mate, one needs many functions in addition to the dedicated "mating" module, including modules of self concept, kinship, justice, monitoring of current well-being, even danger, as well as the intuitive psychology module itself. To the extent that information processing in each of these domains has its own design features, its own particular language and logic, its own way of learning, how can these multiple fields of information be integrated in pursuit of the individual's goals? It is striking that with all the attention currently being paid to the multiplicity of systems from so many perspectives, there is, to my knowledge, so little attention paid to the means by which the systems may be integrated. (I am of course referring to integration on the psychological or functional, rather than neurological level.)

Within many fields, notably the verbal learning approach in psychology, and including fields within cognitive science, philosophy and linguistics, as well as psychoanalysis, there has been a widespread assumption—I would say a shared illusion—that language somehow operates to provide a common integrative device, organizing and regulating the functions of mind; of course this cannot be the case. The interconnection of systems must occur outside of language; the multiple formats exist in all species, as in humans from the beginning of life, well before language is acquired; and the adaptive task of integration of systems must be carried out by all organisms. On a more basic epistemological level, applying to humans after the acquisition of language as in other beings, there must be organization of the experiential field prior to and independent of language, in order to have organized discrete entities to which the verbal labels can be attached.

Far from verbal language operating to solve the problem of integration of diverse representational and processing systems, the problem of interconnection of systems is actually most obvious and acute in the human species for the language module. Human verbal language has specific design features that differ from other forms of human mentation and communication, and also differ from the types of communication and representational forms present in other species. As CHOMSKY, PINKER and others have shown, the language system enables production of an infinite variety of meaningful expressions by arranging and rearranging a finite number of discrete elements in different orders and combinations. With the possible exception of dolphins, who show some features of a generative symbolic system, there is no other species that develops this form of communication system, either naturally or even when subjected to intensive training. The valiant, but doomed campaigns in the 1960s and 1970s to teach language to baby primates served to demonstrate this point. With all the effort and all the techniques that were applied and all the formats that were attempted, the highly intelligent animals could not acquire anything like the communication formats that human children typically learn easily, quickly and without specific instructions.

If it is the case, as PINKER argues, that verbal language evolved in humans as a biological adaptation to communicate information, then the information that needs to be communicated presumably includes everything processed in all the other modules. The specific problem that we will be concerned with here is how to use an instrument with the specific and unique design of the language module to connect to multiple other modules with entirely different designs—"each keyed to the peculiar logic and laws of one domain"; "each with provisions to learn in its own way" (PINKER 1994, p410)

There is a surprisingly sparse literature on the integration of linguistic, cognitive and perceptual processing systems in any of the areas where one might expect to find this point addressed—in developmental psychology, psycholinguistics, cognitive psychology, and also as far as I know in evolutionary psychology.

Cognitive psychologists have generally dealt with this question by ignoring it, denying multiplicity of systems at the basic processing level. From the perspective of the classic symbolic systems, such as propositional or common code models, there is no problem of integration of long-term memory systems; there is only one abstract amodal system to which all verbal and nonverbal input refers and in which all meaning is represented.

The connectionist or PDP approach is built on the premise of separate systems, but no viable integrative models have been produced within this approach. The PDP models developed thus far are restricted to specific tasks. Each unit forms its own pattern of activity through learning, and the activity pattern learned by one network is generally indecipherable to another.

The dual code approach, as formulated by PAIVIO and his colleagues, (PAIVIO 1971, 1986), has addressed the problem of connecting nonverbal and verbal representation to some degree, but accounts only for nonverbal information registered in symbolic forms, not incorporating the problems of integration raised by the type of information processing handled in PDP systems.

Psychoanalysis does address the integration of systems, but generally accounts for this by processes of dominance or absorption, rather than connection among equal components—something like the unification of Berlin. The goal of psychoanalytic treatment is dominance of irrational or infantile systems by presumably more mature and rational ones: making the unconscious conscious; where id was there ego shall be; primary process nonverbal representations giving way to secondary process verbal modes.

To the (minimal) extent that developmental theories do address the question of integration of diverse systems, similar processes of dominance rather than connection are assumed. Thus in PIAGET's theory (1950), sensory motor gives way to concrete and then to formal operations. For BRUNER (1966), sensory and iconic processing give way to symbolic modes.

The pioneering work of the Russian psychologist VYGOTSKY (1986) recognized the separate lines of development of thought and speech, in human infants, as in species other than man, and provided a model for their integration in the second year of human life, in the complex psychological unit of "word meaning". While VYGOTSKY made a major contribution in investigating cognitive life outside of language, and vocalization as distinct from cognition, his integrative format of "word meanings" is verbal symbolic; he remains a verbal mediation theorist, a product of his culture and times. Like PIAGET and BRUNER, although perhaps for different reasons, he leaves out of account the continuation of separate nonverbal and verbal functions throughout life.

An implication of VYGOTSKY's verbal mediation formulation is that he explicitly excludes the domain of emotional expression from the integrative union of vocalization and thought:

"The higher, specifically human forms of psychological communication are possible because man's reflection of reality is carried out in generalized concepts. In the sphere of the emotions, where sensation and affect reign, neither understanding nor real communication is possible, but only affective contagion" (VYGOTSKY 1986, p8).

Thus VYGOTSKY relegates the expression of emotion to the preintellectual roots of speech, differentiated from thought, and restricts the powerful notion of developing word meanings to more neutral cognitive domains. This is particularly surprising in a man who was certainly familiar with CHEKHOV, DOSTOYEVSKY and TOLSTOY. Only WERNER, of the major developmental theorists, has incorporated the possibility of dual lines of cognitive processing, reflecting different functions, for some individuals, throughout life, but again without explicitly addressing the issue of the integration of the separate processing tracks. (WERNER/KAPLAN 1984).

In the remainder of this paper, I will present the multiple code theory of the organization of human information processing that I have developed, which proposes a classification system that recognizes natural divisions of the human psychic terrain; and the corollary theory of the referential process, the means by which the diverse systems may be connected.

The Multiple Code Theory

The multiple systems include two basic formats: the symbolic and what I call the subsymbolic (or nonsymbolic) codes; the distinction between verbal and nonverbal processing is subordinate to this[1]. The differences in processing format underly many of the dimensions of function or content or level of awareness that have been outlined above.

1. Nonverbal and verbal symbolic codes

Symbols—in the sense used here, the information processing or semiotic, not the literary or psychoanalytic sense—are discrete entities that refer to other entities and have the capacity of being combined to make an essentially infinite variety of forms; symbols are familiar to us as images and words. Language is the quintessential symbolic format; from the limited discrete set of phonemes, the infinite varieties of verbal expression are constructed.

Words also have the specific and crucial properties, distinguishing them from nonverbal symbols, of being amodal, arbitrary and abstract, bearing no depictive relationship to the entities to which they refer, with few exceptions, as in onomatopoeia. The detachment from concrete forms gives language its power as a logical processor, ranging across modalities and contents, but the detachment also limits the capacity of language to express nonverbal experience, as VYGOTSKY emphasized. With all its power, language also has the restriction of being a single channel sequential processor; we speak or understand only one verbal message at a time.

Nonverbal images, like words, are digital and discrete, capable of being combined in an essentially infinite variety of new forms. A police artist puts together composites of features according to a witness's description, to approximate a suspect's face. Images differ from words in that they are modality specific and representational, not abstract and arbitrary; images resemble the entities they represent.

Images may occur in all sensory modalities. We are quite familiar with images in the visual system, but also have auditory images, as when a melody or a voice is represented in the mind. While most people may attend primarily to visual and auditory images, others have highly developed imagery of touch or taste or smell. Helen KELLER vividly and evocatively described such images and their central importance in her mental and emotional life.

"A tangible object passes complete into my brain with the warmth of life upon it, and occupies the same place that it does in space; for without egotism, the mind is as large as the universe. When I think of hills, I think of the upward strength I tread upon. When water is the object of my thought, I feel the cool shock of the plunge and the quick yielding of the waves that crisp and curl and ripple about my body. The pleasing changes of rough and smooth, pliant and rigid, curved and straight in the back and branches of a tree give the truth to my hand. … The bulge of a watermelon and the puffed-up rotundities of squashes that sprout, bud and ripen in that strange garden planted somewhere behind my finger-tips are the ludicrous in my tactual memory and imagination" (KELLER 1908, pp10–11).

The generative power of the symbolizing process also applies in all modalities: "My fingers cannot, of course, get the impression of a large whole at a glance; but I feel the parts and my mind puts them together" (KELLER 1908, p12).

In contrast to words, imagery has the power of parallel representation in multiple modalities: we can listen to music, look at the sunset, retrieve a childhood memory and taste wine at the same time—but have trouble talking to our companion if we are listening to the conversation at the next table.

2. The subsymbolic system

Subsymbolic processing is systematic processing that occurs in motoric, visceral, sensory form, as sounds, smells, feelings of many different sorts. (The term "subsymbolic" is taken from its usage in cognitive science, not intended to imply a lower or earlier form.) In cognitive science, such processing is characterized as formally "analogic", processed as very fine variations on continuous dimensions, rather than generated from discrete elements.

Subsymbolic "computations" underly hundreds of common actions from recognizing a familiar voice to entering a lane of traffic, and account as well for complex skills in athletics, and for creative work in sciences and the arts. People in all types of creative fields—painters, sculptors, musicians, dancers, geometers, physicists and many others—operate in highly complex, systematic and differentiated ways in the subsymbolic mode. For all of us, subsymbolic processing accounts for knowing one's own bodily states and responding to the facial and bodily expressions of others, without being able to measure them in discrete units or to categorize them in symbolic form. These are systematic forms of thought with their own processing rules, different from symbolic processing, but equally complex and organized, and equally worthy to be recognized as "thought".

We know this processing as intuition, the wisdom of the body and in other related ways. The patient and analyst communicate profoundly in this mode. REIK's (1964) concept of "listening with the third ear" draws heavily on subsymbolic communication, as I have discussed in detail elsewhere (BUCCI 2001). Many of the modules of intuitive knowledge identified by PINKER operate primarily in the subsymbolic mode, and it is likely that we share some of these with other species.

Subsymbolic processing is modeled in cognitive science by connectionist or parallel distributed processing (PDP) systems (RUMELHART et al. 1986; MCCLELLAND et al. 1986), with the features of dynamical systems, as I have outlined elsewhere (BUCCI 1997)[2]. There is support in neurological observations for the operation of this format. PANKSEPP (1999) refers to global state processes of the brain that are embodied and fundamentally analog—not able to be simulated by digital algorithms. As PANKSEPP says, the models that can handle the full complexity of emotions in the brain will require dynamical systems approaches that account for such analogic processes.

In operating without explicit intention or direction, subsymbolic processes and representations may not be recognized and acknowledged, and may be experienced as in a sense "outside of oneself", outside of the domain of the self over which one has intentional control; what are known as automatic or unconscious processes are likely to be subsymbolic in format. The distinction between symbolic and subsymbolic processes corresponds closely to the continuous vs. categorical distinction and also corresponds in part to the declarative–nondeclarative distinction but is largely orthogonal to the explicit–implicit classification. All types of stimuli, both verbal and nonverbal, have been used in the priming experiments through which implicit memory has been demonstrated.

I argue that the dimension of processing format—subsymbolic vs. symbolic—is more fundamental for mental functioning than the dimension of level of awareness. This applies in psychoanalytic theory as well. Much of what has been categorized as unconscious thought and unconscious communication is really taking place within awareness, but outside of the verbal and often outside of the symbolic zone. The dancer and musician and athlete focus intensely on their performance, not breaking down the functions to elements (which would interfere with performance) but retaining focus on the processes by which the functions are regulated and directed. The regulation of this focus is intentional, but indirect, and as yet not well understood. The tennis player is taught to "keep your eye on the ball"; the musician or actor or dancer maintains focus on particular zones of experience within themselves. If they lose this intense focus, their performance suffers. The operation of different levels of consciousness, and the complex relationship of attention to consciousness are relevant to this type of processing, and need to be further explored (DAMASIO 1994, 1999; LAMBIE/MARCEL 2002; NACCACHE/BLANDIN/DEHAENE 2002). Humans are generally not accustomed to recognizing and regulating their subsymbolic processing; we need to move out of the power of language and also out of symbolic modes to attend to processing in the subsymbolic zones, and to develop powers of regulating these functions. (Meditation has been seen as one way of connecting directly to this unfamiliar processing zone.)

3. The referential process

The design problem in the human information processing system may now be stated in a new way: how to connect the analogic processes of the subsymbolic system, processed in parallel in multiple modalities and experienced to some degree as outside of oneself, to the arbitrary, abstract, discrete elements of the single channel verbal code. How do we arrange discrete elements (lexical items) in particular orders and combinations, to express the nuances of emotion; to describe movements, which are continuous and complex; to represent the analogic spatial relations of geometry and physics; or to communicate any of the other myriad forms of human knowledge and experience processed in the subsymbolic format.

In solving the problem of integration of diverse processing formats, it is imagery—symbolic nonverbal representations—that plays the pivotal role. As nonarbitrary representational elements in sensory modalities with synchronous multi-modal presentation, images connect to the subsymbolic mode; as discrete representational elements (nonverbal symbols), which can be composed of component elements, simulated by digital algorithms, they connect to the symbols of the verbal code.

The referential process is the process of connecting diverse forms of nonverbal experience to one another and to words. I have identified four phases of the referential process; we may see these playing out in many domains—in creative work in science and the arts, as well as in the psychoanalyst's office.

4. Phases of the referential process

a. Subsymbolic. This is a phase of exploration, involving arousal of a cognitive or emotional schema, or component of a schema, which is not yet in symbolic form, and which may never have been verbalized or even symbolized as imagery. In activation and expression of an emotion schema, the affective core of somatic and sensory experience will be dominant in this phase. The scientist or mathematician may have access to spatial or other perceptual representations in seemingly vague and shifting analogic forms.

b. Early symbolic. This is a transitional phase in which the analogic is made discrete. The individual may focus on fragments of imagery or specific bodily experiences, which have been only minimally verbalized. The experience may be verbalized as somatic complaints or descriptions of images, or in some cases represented as symbolic pictorial representations or enactments.

c. Symbolic narrative phase. This phase is characterized by retrieval or construction of an organized image or idea, which can be reported in verbal form, as a narrative of a memory, fantasy, dream, or episode of current life. In communication of an emotional experience, the speaker is here connecting the subsymbolic processes of the affective core to discrete images of the people, places and objects that populate the emotion schemas. The narrative constitutes a metaphor of the activated emotion schema, carrying across the emotional meaning in discourse, in many cases without the meaning being understood in symbolic form, by speaker or listener.

d. Reflection. This involves elaboration of the meaning of the material reported in the narrative phase. The emotion that is reported in narrative form is explicated and recognized in this phase. Optimally, in creative work in sciences and the arts, as well as in understanding of one's emotional life, this phase will lead to development of new categories, arousal of new schemas or components of schemas, and recursion of the cycle on a deeper level.

Evidence for Imagery as the Pivotal Mode

There is considerable converging evidence from many sources that support and elaborate the functioning of prototypic imagery as pivotal in the transition from subsymbolic to symbolic forms. I can mention a few of these sources only briefly here; for further detail, see BUCCI (1997).

KOSSLYN (1987) introduced the concept of functionally equivalent classes of stimuli as linking continuous and categorical subsystems. The categorical processor of the visual system ignores variation within ranges that are experienced by the organism as functionally equivalent, responding to this range of manifestations as if they were the same. The functionally equivalent classes of stimuli—the form of a smile on different faces; the persona of mother in her different manifestations—are represented as the prototypic images that populate our memory schemas. While KOSSLYN's formulation of the construction of prototypic images was developed specifically for the visual modality, his basic formulation may be extended to account for construction of prototypic imagery in other sensory modalities, in ways determined by their specific features, and to account for cross-modal imagery as well. In addition to representations of objects, the categorical system also has the power to represent relations that are prototypic, such as being *over*, *under* or *next to* something. Such relationships among objects, like objects themselves, are then represented in memory in prototypic form, which can be translated into words.

ROSCH's (1978) notion of the "fuzzy" categorization of prototypes, like KOSSLYN's notion of functionally equivalent classes, provides a basis for chunking nonverbal information into discrete symbolic representations. ROSCH's system incorporates prototypes built on functional as well as perceptual equivalence; for example, prototypic representations of an object to sit on, or an object with legs and a surface on which we place other things. ROSCH has also applied her concept of natural prototypic structure to the organization of the flow of experience in terms of episodes or events; thus her formulation accounts for event structures as themselves having prototypic form.

STERN (1985) has proposed a similar formulation of how the components of the child's emotional life and emotional memory are organized through development of prototypic memories, which he has termed Representations of Interactions that have been Generalized (RIGs). Repeated episodes incorporating actions, experience in all sensory modalities, and visceral experience, and involving other people in relationship to oneself, form the prototypic memory structures, which provide order to experience, including emotional experience, and which are the basis for the development of a sense of self.

MANDLER (1992) proposes a similar notion in her formulation of image-schemas, and has also addressed the means by which perceptual experience is mapped onto language. MANDLER's theory of perceptual analysis is also compatible with the notion of image-schemas as formulated by cognitive linguists (e.g., JOHNSON 1987; LAKOFF 1987; LANGACKER 1987). Image schemas consist of mappings from spatial structures, which include general aspects of the trajectories of objects and their interactions in space, onto conceptual and linguistic structures that represent the meanings of objects, and the kind of events they participate in.

All of the formulations that have been outlined here, from several different perspectives, support the notion of organization occurring within the nonverbal system, connecting the global analogic formats that are distant from language, and that may be manifest consciously or be out of awareness, to categorical imagistic entities—prototypic imagery of many types—that are then capable of being mapped onto the discrete, symbolic elements of the verbal code[3]. My empirical work has been focused on developing measures of each phase of the referential process, applied to many types of discourse, from poetry, memoirs, and political speeches to psychoanalytic sessions (BUCCI 1997; MASKIT/BUCCI/ROUSSOS in preparation).

Illustrations of the Referential Process

The processes of creative work in science and the arts, like emotional exploration in psychoanalysis, are best characterized, not as making the unconscious conscious, but as connecting subsymbolic to symbolic representational modes. This requires categorizing or chunking the continuous flow of subsymbolic experience into discrete units that are valid and meaningful as categorizing the underlying experiential domain. We do not simply retrieve stored representations and bring them to consciousness; the process involves construction of new forms. This is a central aspect of communication that takes place whenever we must find the words, often in metaphoric forms, to express sensory or somatic or motoric experience—how something tastes or looks or feels, or how to execute an athletic or dance movement. For people whose business is creative exploration, the process may be more elaborate and in some cases more accessible to introspection.

In his essay "Mathematical Creation", POINCARÉ gives several vivid examples of his process of discovery, as in the following brief description, which may be outlined in terms of the stages of the referential process. POINCARÉ writes:

"For fifteen days I strove to prove that there could not be any functions like those I have since called FUCHSIAN functions. I was then very ignorant; every day I seated myself at my work table, stayed an hour or two, tried a great number of combinations and reached no results. One evening, contrary to my custom, I drank black coffee and could not sleep."

This would be characterized as a phase of arousal, leading then to early or transitional symbolic processing: "Ideas rose in crowds: I felt them collide," and then to the symbolic mode: "until pairs interlocked, so to speak, making a stable combination. By the next morning I had established the existence of a class of FUCHSIAN functions, those which come from the hypergeometric series."

Finally he reaches a phase of reflection and integration, which is simple and straightforward in this case: "I had only to write out the results, which took but a few hours" (POINCARÉ, in NEWMAN 1956, p2044[4]).

The process of discovery is cyclical or recursive; repeating throughout POINCARÉ's work. Once the subsymbolic search has begun it continues to a large extent outside of awareness and may be facilitated by alternately withdrawing attention and returning to the search. In POINCARÉ's descriptions, many of his major discoveries were prefaced by some version of *turning away*—a trip to the seaside, entering military service, a geological excursion—and the insights appeared as if from outside himself, as I have discussed in more detail elsewhere (BUCCI 1997).

Similar patterns of discovery may be seen in different forms in the work of other scientists and mathematicians, as well as writers. In his studies of creative thought, based on extensive reports by leading scientist and mathematicians, HADAMARD (1954) identified four stages in the process of discovery, which he termed *Preparation, Incubation, Illumination* and *Verification,* and which correspond closely to the phases of the referential process identified here. HADAMARD gives many examples of this process, based on introspective reports of leading scientists and mathematicians, including POINCARÉ and others.

EINSTEIN seemed to have considerable access to his cognitive operations as he moved from what I would characterized as the subsymbolic to symbolic modes:

"The psychical entities which seem to serve as elements in thought are certain signs and more or less clear images which can be "voluntarily" reproduced and combined …"

"[T]his combinatory play seems to be the essential feature in productive thought—before there is any connection with logical construction in words or other kinds of signs which can be communicated to others…

"The above mentioned elements are, in my case, of visual and some of muscular type.

"Conventional words or other signs have to be sought for laboriously only in a secondary stage, when the mentioned associative play is sufficiently established and can be reproduced at will" (EINSTEIN, quoted in HADAMARD 1954, pp142–143).

The stages of the referential process from arousal to reflection may also be seen in specific and elaborated form in the creative process of recovery and reconstruction of memories as described most vividly by PROUST (1934). I will illustrate this progression with the famous passage of the tea and madeleine that is part of the overture to *Swann's Way* and that opens PROUST's entire six volume project.

a) First the subsymbolic phase, arousal of the memory schema; PROUST's mother offers him some tea, which he does not ordinarily drink:

"No sooner had the warm liquid, and the crumbs with it, touched my palate than a shudder ran through my whole body, and I stopped, intent upon the extraordinary changes that were taking place. An exquisite pleasure had invaded my senses, but individual, detached, with no suggestion of its origin. … Whence could it have come to me, this all-powerful joy? … What did it signify? How could I seize upon and define it?…

"Seek? More than that: create. It [my mind] is face to face with something which does not so far exist, to which it alone can give reality and substance, which it alone can bring into the light of day" (p34).

b) PROUST then provides an extraordinary description of the early symbolic transitional phase:

"…I feel something start within me, something that leaves its resting-place and attempts to rise, something that has been embedded like an anchor at a great depth; I do not know yet what it is, but I can feel it mounting slowly; I can measure the resistance, I can hear the echo of great spaces traversed.

"Undoubtedly what is thus palpitating in the depths of my being must be the image, the visual memory which, being linked to that taste, has tried to follow it into my conscious mind. But its struggles are too far off, too much confused; scarcely can I perceive the colourless reflection in which are blended the uncapturable whirling medley of radiant hues, and I cannot distinguish its form" (p35).

That is what it feels like as the analogic is becoming discrete.

(c) The central metaphor of the memory schema then emerges in the symbolic phase. After many struggles, turning away in exhaustion, then returning to the search:

"…suddenly the memory returns. The taste was that of the little crumb of madeleine which on Sunday mornings at Combray …. my aunt Léonie used to give me, dipping it first in her own cup of real or of lime-flower tea" (p36).

The sight of the madeleine had not been sufficient to awaken the memory; the visual appearance of the familiar object had lost the power of expansion, which the unexpected sensory experience of taste, the cake dipped in tea that PROUST did not ordinarily drink, retained:

"But when from a long-distant past nothing subsists, after the people are dead, after the things are broken and scattered, still, alone, more fragile, but with more vitality, more unsubstantial, more persistent, more faithful, the smell and taste of things remain poised a long time, like souls, ready to remind us, waiting and hoping for their moment, amid the ruins of all the rest; and bear unfaltering, in the tiny and almost impalpable drop of their essence, the vast structure of recollection" (p36).

(d) Reflection, elaboration, expansion. The taste provides the essential connection to the exquisite pleasure, the all-powerful joy that PROUST seeks. Once he has recognized the taste, he has access to imagery of the places and the people of that past time. But the emotional meanings remained to be found: "I did not yet know and must long postpone the discovery of why this memory made me so happy". As the mathematician verifies at his leisure the results that appear to him in the moment of illumination, PROUST goes on to seek the emotional meanings embedded in the images of Combray. The process is recursive; new elements of the memory schemas emerge, evoking new explorations and new constructions. PROUST goes on to give substance and reality to the memories, to elaborate the structure of his recollection, in the six volumes of his monumental work.

We can see this process occurring for each of us, as we struggle to construct and retrieve and communicate inner experience of all types, without losing its essence. The great human advance was not so much to develop a vocal system with a finite number of discrete elements capable of being combined to create an infinite number of linguistic forms—but to develop a system in which these linguistic forms presumably have the capability to represent everything else we know, in its multiple diverse modes. We now are beginning to recognize that we are only partly there in achieving this advance; it may be useful for us to recognize the intrinsic nature of the gaps in the human information processing system, operating to varying degrees for all of us, not just in special diagnosed conditions, and to recognize the implications of these gaps for the human condition. We need to use the special human capacity for self-observation and self-regulation to develop means of connecting systems within oneself, and to see how this provides the infrastructure that enables us to connect to others.

Author's address

Wilma Bucci, Derner Institute of Advanced Psychological Studies, Adelphi University, Garden City NY 11530, USA.
Email: wbucci@adelphi.edu

Notes

1 The possibility of a system that would be characterized as "verbal subsymbolic" remains open to question. Paralinguistic aspects of language including pausing rhythms and intonation patterns, and aspects of the sound of speech, e.g., as in onomatopoeia or more generally in poetry, may operate on a subsymbolic level, as may emotional vocalizations (e.g., sighing, giggling), but the words of language themselves appear to be intrinsically digital and discrete elements. The operation of a verbal subsymbolic code needs to be further explored.

2 To review this briefly: A connectionist system is a network with a set of interconnected nodes. The theoretical connectionist networks are designed to model the structure of neural networks, but are simpler than the actual physical (brain) systems, and retain the status of psychological (mental) models. At any given time, each node is in a particular state of arousal, and the state of the system is dependent on the level of arousal of each of the nodes, and on the patterns of interconnections and weights; these are determined in large part by learning and experience. The pattern of interconnections and matrix of weights may be described as a dynamical system. Given the input of particular states of arousal for each node, the system of connections and weights assigns new states of arousal for each, which then function as new input. The system continues this iterative process, testing the match with the desired target (entering a busy highway, turning a sail, recognizing a face or voice or wine) and evaluating the error, the difference between the actual and desired state, until the error is small enough to accept the match.

3 The phases of the referential process may also be seen as related to DAMASIO's (1994, 1999) concept of expanding levels of consciousness as associated with emotion and the sense of self, as I have discussed in detail elsewhere (BUCCI in press).

4 Lecture delivered before the Psychological Society in Paris, published in NEWMAN (1956), pp2041–2050.

References

Bornstein, M. (1979) Perceptual development: Stability and change in feature perception. In: Bornstein, M./Kessen, W. (eds) Psychological development from infancy: Image to intention. Lawrence Erlbaum Associates: Hillsdale NJ, pp. 37–81.

Bruner, J. S. (1966) On cognitive growth. In: Bruner, J. S. et al. (eds) Studies in cognitive growth. John Wiley: New York.

Bucci, W. (1997) Psychoanalysis and cognitive science: A multiple code theory. Guilford Press: New York.

Bucci, W. (2001) Pathways of emotional communication. Psychoanalytic Inquiry 20:40–70

Bucci, W. (In press) The referential process, consciousness, and the sense of self. Psychoanalytic Inquiry.

Cohen, N. J. (1984) Preserved learning capacity in amnesia: Evidence for multiple memory systems. In: Squire, L. R./ Butters, N. (eds) Neuropsychology of memory. Guilford Press: New York, pp. 83–103.

Damasio, A. R. (1994) Descartes' error: Emotion, reason and the human brain. Avon Books Inc: New York.

Damasio, A. R. (1999) The feeling of what happens. Harcourt Brace & Co: New York.

Farah, M. J. (1984) The neurological basis of mental imagery: A componential analysis. Cognition 18:245–272.

Farah, M. J./McClelland, J. L. (1991) A computational model of semantic memory impairment: Modality specificity and emergent category specificity. Journal of Experimental Psychology–General 120:339–357.

Fodor, J. A. (1983) Modularity of mind. MIT Press: Cambridge MA.

Freud, S. (1964) An outline of psycho-analysis. In: Strachey, J. et al. (eds) The standard edition of the complete psychological works of Sigmund Freud, Volume 23. Hogart Press: London, pp. 139–207. originally published in 1940.

Gallistel, C. R. (1990) The organization of learning. MIT Press: Cambridge MA.

Gardner, H. (1983) Frames of mind: The theory of multiple intelligences. Basic Books: New York.

Gazzaniga, M. S. (1985) The social brain. Basic Books Inc: New York.

Gazzaniga, M. S. (1988) The dynamics of cerebral specialization and modular interactions. In: Weiskrantz, L. (ed) Thought without language. Clarendon Press: Oxford, pp. 430–450.

Hadamard, J. (1954) The psychology of invention in the mathematical field. Dover Publications: New York. Originally published in 1945.

Johnson, M. (1987) The body in the mind: The bodily basis of meaning, imagination, and reasoning. University of Chicago Press: Chicago.

Keil, F. (1989) Concepts, kinds, and conceptual development. MIT Press: Cambridge MA.

Keller, H. (1908) The world I live in. The Century Co: New York.

Kosslyn, S. M. (1987) Seeing and imagining in the cerebral hemispheres: A computational approach. Psychological Review 94:148–175.

Lakoff, G. (1987) Women, fire, and dangerous things: What categories reveal about the mind. University of Chicago Press: Chicago.

Lambie, J. A./Marcel, A. J. (2002) Consciousness and the varieties of emotional experience: A theoretical framework. Psychological Review 109(2):219–259.

Langacker, R. (1987) Foundations of cognitive grammar (Vol. 1). Stanford University Press: Stanford CA.

McClelland, J. L./Rumelhart, D. E./The PDP Research Group (Eds.) (1986) Parallel distributed processing: explorations in the microstructure of cognition, Vol. 2 Psychological and biological models. MIT Press: Cambridge MA.

Mandler, J. (1992) How to build a baby II: Conceptual primitives. Psychological Review 99:587–604.

Maskit, B./Bucci, W./Roussos, A. (2002, In preparationAL-READY APPEARED?**)** Capturing the flow of verbal interaction; The Discourse Attributes Analysis Program.

Minsky, M. (1975) A framework for representing knowledge. In: Winston, P. H. (ed) The psychology of computer vision. McGraw-Hill: New York, pp. 211–277.

Naccache, L./Blandin, E./Dehaene, S. (2002) Unconscious masked priming depends on temporal attention. Psychological Science 13:416–424.

Newman, J. R. (1956) The world of mathematics 4. Simon and Schuster: New York.

Paivio, A. (1971) Imagery and verbal processes. Holt, Rinehart & Winston: New York.

Paivio, A. (1986) Mental representations: A dual coding approach. Oxford University Press: New York.

Panksepp, J. (1999) Emotions as viewed by psychoanalysis and neuroscience: An exercise in consilience. Neuro-Psychoanalysis 1:15–38.

Piaget, J. (1950) The psychology of intelligence. Routledge & Kegen Paul: London.

Pinker, S. (1994) The language instinct. HarperPerennial: New York.

Posner, M. I./Snyder, C. R. R. (1975) Attention and cognitive control. In: Solso, R. (ed) Information processing and cognition: The Loyola symposium. Erlbaum: Hillsdale NJ, pp. 55–85.

Proust, M. (1934) Remembrance of things past (Translated by C. K. Scott Moncrieff). Random House: New York.

Reik, T. (1964) Listening with the third ear: The inner experience of a psychoanalyst. Pyramid Books: New York. Originally published in 1948.

Rosch, E. (1978) Principles of categorization. In: Rosch, E./ Lloyd, B. B. (eds) Cognition and categorization. Lawrence Erlbaum: Hillsdale NJ, pp. 27–48.

Rumelhart, D. E./McClelland, J. L./The PDP Research Group (1986) Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundations. MIT Press: Cambridge MA.

Schacter, D. L. (1989) Memory. In: Posner, M. A. (ed) Foundations of cognitive science. MIT Press: Cambridge MA, pp. 683–725.

Squire, L. R. (1992) Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. Psychological Review 99:195–231.

Stern, D. N. (1985) The interpersonal world of the infant. Basic Books: New York.

Tulving, E. (1972) Episodic and semantic memory. In: Tulving, E./Donaldson, W. (eds) Organization of memory. Academic Press: New York, pp. 381–403.

Tulving, E. (1985) How many memory systems are there? American Psychologist 40:385–398.

Vygotsky, L. (1986) Thought and language. MIT Press: Cambridge MA. Originally published in 1934.

Werner, H./Kaplan, B. (1984) Symbol formation. Lawrence Erlbaum: Hillsdale NJ.

Robert E. Haskell

The New Cognitive Unconscious

A Logico-Mathematic-Structural (LMS) Methodology and Theoretical Bases for Sub-Literal ($S_{ub}L_{it}$) Cognition and Language

Abstract

A novel cognitive and psycho-linguistic methodology developed for the analysis and validation of unconscious cognition and semantic references found in ostensibly literal verbal narratives is empirically demonstrated. There has been neither a systematic cognitive theory nor a methodology for recognizing, analyzing, and validating unconscious phenomena. The lack of systematic recognition and methodology leaves such "interpretive" meanings open to clear criticism as to whether unconscious linguistic reference in fact exists. Unconscious reference in the verbal narratives is reconceptualized as sub-literal ($S_{ub}L_{it}$) reference. It is suggested that the integrally systemic, structural, and internally consistent set of operations constitute a new view of unconscious cognitive processing. Issues of experimentation, pertinent knowledge-base, and theoretically supportive findings from animal and infant research are presented. Finally, a biological evolutionary framework is briefly presented to explain the origin and development of $S_{ub}L_{it}$ cognition and reference

Key words

Unconscious, narrative, cognition, linguistic, language, methodology.

Recognition of unconscious referents in spoken language has been ubiquitous for centuries. Essentially, the hypothesis has been that some literal words, phrases, and topics in oral (and occasionally written) communications denote consciously unintended referents. Perhaps the most common and simple examples being puns and double entendres. Despite the widespread perception of unconscious reference in verbal narratives, analysis and validation remain intuitive, if not serendipitous, and therefore problematic in terms of veridicality. As developed from a cognitive and psycho-linguistic methodology (see Haskell 1991, 2003a)[1] Haskell (1983) has reconceptualized unconscious semantic references expressed in spoken language as sub-literal ($S_{ub}L_{it}$). $S_{ub}L_{it}$ reference is defined as the presence of unconscious semantics coexistent with the literal or conventional meaning of words.[2]

Briefly, it has been recognized that in social discussions where a concern of discussants is about an *authority figure* present in a conversation that increased literal references to topics about *God, heads of state, police, parental* and other *authority figures* appear in the narrative. Speakers are not aware of the dual semantic structure of the literal topics reflecting their concerns about the authority figure. Historically, such topics have been seen as metaphor-like expressions of unconscious concerns that discussants have about the authority figure.[3] A $S_{ub}L_{it}$ narrative, then, is a literal story, sentence, or phrase that exhibits a coexistent meaning or referent. To varying degrees such phenomena have been conceptualized in psychoanalytic and other domain frameworks, but not in cognitive and linguistic terms. In addition, there has been neither a systematic theory nor a methodology for recognizing, analyzing, and validating $S_{ub}L_{it}$ phenomena. The lack of systematic recognition and methodology left such "interpretive" meanings open to justified criticism as to whether unconscious meaning, in fact, exists.

In addition to developing a systemic methodology where none previously existed, there have been two problems associated with $S_{ub}L_{it}$ findings. The first has been to base the findings firmly within a cognitive scientific framework and not have them relegated to the domains of psychoanalytic, discourse analysis, or literary theory as the tendency has been to do with material involving complex linguistic analysis and unconscious referents.[4] The findings are too cognitively significant to be relegated to these domains.

The second problem has been to develop a method which includes complex "meaning" or semantics not just an abstract architecture as in most of linguistics and cognitive science, e.g., with classic speech-error and slips- of-the-tongue research (e.g., BAARS et al. 1992; FROMKIN 1973; NORMAN 1981). While $S_{ub}L_{it}$ language appears to utilize lawful mechanisms identified in linguistics and in the analysis of speech-error research, $S_{ub}L_{it}$ phenomena demonstrate a class of "errors" and "slips" possessing underlying intentionality and co-existent meaning. Many of the cognitive and linguistic operations demonstrated in $S_{ub}L_{it}$ methodology (HASKELL 1991, 2003a) constitute a new view of unconscious cognitive processing.[5]

This paper demonstrates some of the novel cognitive operations involved in $S_{ub}L_{it}$ meaning utilizing literal numerical references contained in a verbal narrative.[6] Just as words can function subliterally, so can numbers in narratives have unconscious co-existent referents (see HASKELL 2003a, 1983). Counterintuitively, using numbers in narratives to illustrate the novel cognitive operations may be seem more controversial than analyzing unconscious semantics of words, phrases, sentences and topics, and may indeed even appear to raise the specter of a New Age occult numerology. Numbers contained in a narrative are used, however, because unlike analyzing the meaning (semantics) of words, numbers found in topics are relatively bounded and concrete and can thus be more clearly and succinctly analyzed as well as systemically tracked throughout the narrative (see, *[12.] Arithmetical*).[7]

In initial support of the veridicality of the $S_{ub}L_{it}$ numerical series to be analyzed below, it should be noted the $S_{ub}L_{it}$ numeric findings have been consistently and repeatedly found in all verbal narratives examined by HASKELL. In addition, HASKELL/BADALAMENTI (in press) have recently found that the series of $S_{ub}L_{it}$ numeric topics analyzed here exhibit an algebraic structure, suggesting not only the veridicality of the findings but also the intriguing possibility that this logico-mathematic structure may subserve cognition.

Method and Procedures

The numerical data analyzed here were generated in a small group dynamic laboratory where fifteen participants engaged in a non-structured discussion about leadership and group level dynamics. The exact composition of the group is contextually important for mapping and analyzing $S_{ub}L_{it}$ numeric referents. The data used in this analysis are from a transcribed protocol of a tape recording of one entire narrative session.[8]

The protocol was subjected to the logico-mathematic-structural (LMS) controlling methodology (HASKELL 1991, 2003a). *Working from a transcribed protocol made it possible to systematically account for each numerical reference mentioned in a topic during the session. This virtually eliminated bias in selecting only certain numerical topics to illustrate a specific $S_{ub}L_{it}$ referent.*

In analyzing the numeric content of the narrative topics below, it is important to delineate the numeric composition of the group. It was composed of 15 members, of whom 12 were present. Members included: 1 male, very active; 11 females, of which 1 was an older woman who was quite active, and 10 females about 18–20 years old, of which 5 were active, making 6 total active females; 7 total members were active, 8 counting the trainer; 2 males counting the trainer; 2 dominant leaders, a male and an older female, 3 counting the significant role of the trainer, with a total membership during the session of 13 counting the trainer. Most members were relative strangers to each other.

The unstructured mode of the discussion that occurs in these circumstances generates conditions of (a) low social structure, (b) high ambiguity, (c) uncertainty, (d) interpersonal conflict involving personal or social taboos or etiquette, (e) free-floating discussion, (f) uncomfortable pauses and silences, and (g) generates affective or emotional arousal. It is theorized that these conditions create an optimal cognitive state in which nonconscious affective and linguistic schemas are activated, thus generating $S_{ub}L_{it}$ topics. $S_{ub}L_{it}$ material, however, also occurs in most everyday verbal discourse situations. The hypothesis is that narrative literal topics and their variations are generated from and correspond to affective schemata of discussants about the dynamics in the actual narrative situation. It is important methodologically to note that repeated interviews outside of the narrative situation to ascertain if discussants were aware of the dual referents of their narrative discourse have invariably invoked surprise, laughter, and incredulity.[9]

Transformations and Permutations of an Affective Schema
- (1) 3 Different Options
- (2) 3 Weeks Ago
- (3) 3 Hours Before
- (4) The 3rd Stream
- (5) 3 Drinks
- (6) On the 3rd Day

Permutations of the General Affective Schema:
- *Semantic: Status and Age Attributions*
  - (7) 3 Lucky Spots Bar
  - (8) 3 Seniors
  - (9) 3 Old Greyhound Buses
- *Numerical:*
  - *Gender and Age Attributions*
    - (10) This 1 Girl Who Was With 2 Guys
    - (11) Under 21 Years of Age
    - (12) John Smith, 21 years old
- *Part/Whole Attributions*
    - (13) 2 or 3 Weeks
    - (14) 3 of the 10 People

Figure 1

Parsing the Triadic Structure of Numeric Series

This exemplification of $S_{ub}L_{it}$ reference is based on two major categories involving numerical cognitive operations (*[12.] Arithmetical Operations; [13.] Logico–mathematic Operations of Representation*). The following fourteen numerical topics containing the number *"3"* are from a larger set of literal topics that occurred in a single verbal narrative session where there were three dominant members.[10]

The titles are the verbatim language used in relating the stories. The affective schemata of the discussants about *3 dominant members* generated the data about a triad composed of (a) an older woman, (b) a slightly older male, and (c) the trainer/researcher who constituted the dominant subgroup. The remaining membership consisted of the10 young females. The set of fourteen stories indicates an affective schema about the *3 dominant members*, as each narrative involves a representation containing the number *3* in some form.[11] Two basic cognitive operations divide the stories into two separate sets. The first is a *transformation operation,* the second a *permutation operation.* Categorized by these two operations, the topics are outlined in Figure 1.

The complete series of fourteen topics constitutes a general inclusive logico–mathematic set or group expressing a basic $S_{ub}L_{it}$ schema about the *3 dominant members* in the discussion; hence narratives containing the number "3".

This general set of fourteen topics is then further divided into two subsets, *transformations* and permutations of that general schema. The first six topics appear to be simple transformations of the basic affective schema about the *3 dominant members*. Transformations, then, are single-leveled repetitions of this schema with each transformation repeating the basic schema. *Permutations*, as indicated by topics 7–14, are complex differentiations of specific aspects or elements of the basic affective schema. In this case the $S_{ub}L_{it}$ schemata express differentiations *within* the *3 dominant member* subgroup.

Topics 7–14 are subsets with one set expressed in a semantic mode, the other in a numerical mode. The first subset of permutation topics, 7–9, constitutes permutations semantically expressed. Topic number 7 about a bar called the */3 Lucky Spots/* reflects a $S_{ub}L_{it}$ affective schema with *position status*, as indicated by the vernacular phrase *lucky spots*. Topics 8 and 9 about */3 Seniors/* and */3 Old Greyhound Buses/* express a $S_{ub}L_{it}$ affective schemata about *age status* (i.e., the 3 dominant members who were older than the rest of the members) as indicated by the terms "seniors", and "old".

The second set of permutations, topics 10–13, is constituted by numerical representations regarding the gender and composition of the 3 dominant members (i.e., 1 female and 2 males). For example, the literal topic of */This 1 Girl* Who was With *Two Guys/* is a $S_{ub}L_{it}$ reference to the exact numerical and gender composition of the 3 dominant members subgroup. In the literal topic of being */under 21 Years of Age/* the number "2" in the reference to */21/* years old, again, sub-literally represents the *2 males,* with the */1/* representing the *1 older woman*. Thus, in $S_{ub}L_{it}$ operations the */21/* actually represents two separate numbers that are then added together, i.e., 2 + 1 = 3. This analysis is partially based on (a) the use of thepreposition *under*, meaning that the remaining discussants were subordinate or under the dominance of the triad, in conjunction with the fact (b) that the topic was literally about being *under age,* referencing that the rest of the discussants were younger than the triad. Similarly, topic 12 about a hypothetical */John Smith, 21 years old/*, also represents members' issue with the age difference as indicated by the literal topic having to do with age.

It is methodologically significant that the topic about */John Smith, 21 years old/* was generated by the male member of the *3 dominant members* who was older and about *21 years old.* In addition, the topic is again a $S_{ub}L_{it}$ numerical equation represent-

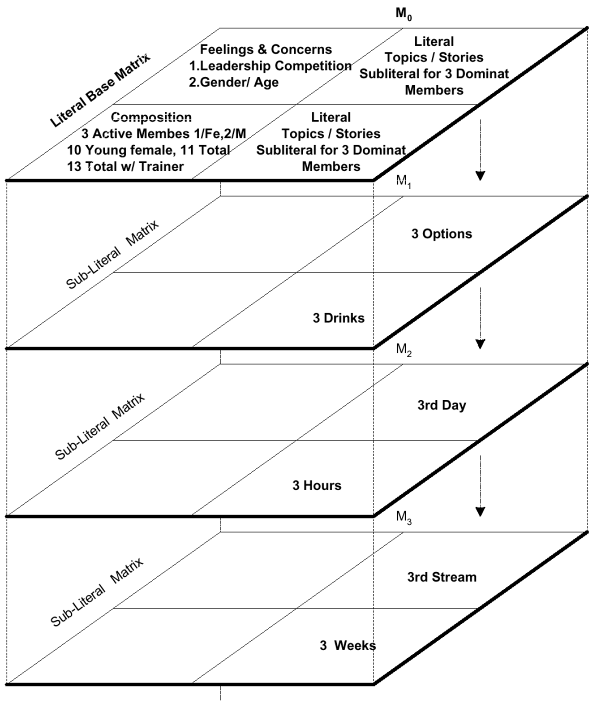

Figure 2. Transformational numerical lattice matrix

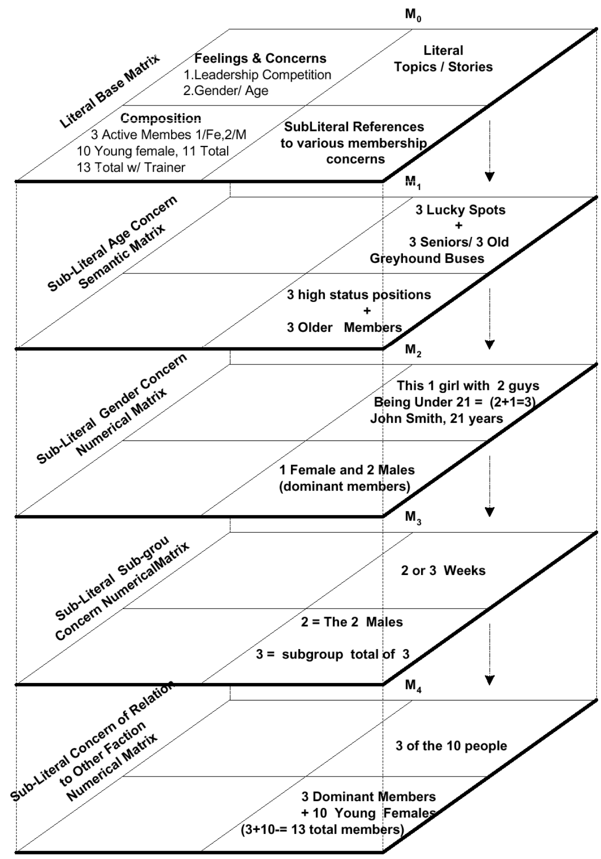

Figure 3. Permutational numerical lattice matrix

ing the *3 dominant members*, i.e., 2 + 1 = 3. Finally, the topic was not associated with the preposition /*under*/ which indicated social hierarchy as was the case in topic 11, but instead the topic was about the age of a male named John Smith. Topic 13, /*2 or 3 Weeks*/, expresses an affective schema about 2 of the dominant members in relations to the total subgroup of 3 (*[12.] Arithmetical Operations; [13.] Logico–mathematic Operations of Representation*).

The third set of permutations exemplified by topic 14 about /*3 of the 10 People*/ expresses an affective schema about a part/whole relationship; the /*3 of the 10 People*/ topic represents a schema about the relationship of the *3 dominant members* to the rest of the group as indicated by the number /*10*/ which represents the remaining *10 members in the discussion*.[12] See Figures 2 and 3.

Because the numerical topics involving"3s" are integrally connected to other numerical topics reflecting additional factions represented within the protocol, this permutation is an access topic into other numeric topics which can not be presented here and will not be further dealt with. It is mentioned to suggest the possibility that there are additional related numerical topics within the protocol that are integral to the 14 topics analyzed.

Summary of Validation Operations of the Numerical Narrative Series

This section will continue with a more detailed focus on the series of narratives containing the number /*3*/ which are hypothesized to be $S_{ub}L_{it}$ topics about the *3 dominant members* and will outline twenty four (a) *Structural Consistencies*, (b), *Linguistic Consistencies* and (c) *Associative Consistencies* that are integral to the above numerical narrative series. These three categories, however, are not necessarily mutually exclusive. A single topic may exemplify more than one of these categories. It is these consistencies found in this series of topics which must be explained; they are not explainable from a literal perspective. These consistencies constitute aspects of the LMS methodology.

Structural consistencies. Structural aspects of the $S_{ub}L_{it}$ topics refer to the isomorphic or identical features of the *literal topics with the actual membership structure* of the conversation. The first four structural consistencies presented below are straightforward literally, requiring no "analysis" to map them to the actual narrative situation.

1. All 14 stories contained the number "3" in one form or another,

2. All 14 stories were methodologically found to structurally correspond to the triadic leadership mapping onto the actual situation of *3 members dominating* the group dynamic (*[1.1.] Matching Operations; [1.2.] Isomorphic Mapping Operations*).

3. With the exception of topics 1, 7, and 12 which are different contextually, the remaining 11 topics were generated only by members who had negative affective schemata about the triadic leadership structure (*[4.1.] Resonance Operations; [4.2.] Sociometric Operations)*.

4. The actual structure of the triad was specifically broken down into its correct $S_{ub}L_{it}$ sub-numerical components in the topics. For example, in topic 10, /*This 1 Girl* Who Was With *2 Guys*/ equates to the 1 female and two male leaders, hence, 1 + 2 = 3. Similarly, numerical topic 11 /*Under 21 Years of Age*/ by an addition operation of the /*2*/ and the /*1*/also equates to 2 + 1 = 3, as does topic 12 /John Smith, *21 years old*/, i.e., 2 + 1 = 3 (*[12.1.] Numerical Representation*).

5. The numeric structure of the literal topic was specifically broken down into its correct $S_{ub}L_{it}$ *gender* components e.g., in topic 10 /This *1 Girl* Who Was With *2 Guys*/ (*[4.3.] Gender Reference Operations*).

6. The composition of the actual triadic structure was correctly and sub-literally differentiated by *age* difference from the rest of the group by term *seniors* in topic 8 about /*3 seniors*/, and by the term *old* being associated with the /*3 old greyhound buses*/ in topic 9, and in topic 12, by /John Smith, *21 years old*/ (*[4.4.] Age Reference Operations*).

7. Adding the numbers in the topic /*3 of the 10 People*/, i.e., *3 + 10*, totals to *13*, the exact membership of the group that session (see just below for this computational operation).

8. The triadic structure was correctly delineated from the rest of the *10 young females* in topic 7, about /*3 of the 10 People*/… who came into a bar. On the literal level the number /*3*/ is included within a total number of /*10*/ (i.e., 3 *of* the 10 people), but in $S_{ub}L_{it}$ terms they are separate, thus adding to 13. The literal arithmetic structure of the number /*3*/ being included as a part of the number /*10*/ would not have fit the total group membership. In other words to have said something like "3 people came in and sat down with the other 7 people at the bar", would have precluded the adding of 3 and 10 to total the 13 members (see, *Mathematical Computation Operations, [8.1.], [8.5.], [9.5.]*).

9. The triadic structure was further broken down into its components. For example, in Topic 6, the reference to the liquor control board men who were coming /Like in *2 or 3* Weeks/, with the /*2*/ corresponding to the 2 males, and the /*3*/ referring to the total triadic structure (see, *Mathematical Computation Operations, [8.1.], [8.5.], [9.5.]*).

10. The remaining membership of the group was correctly differentiated from the triadic structure by a literal description of the people in the bar: /*Over half of them were under age*/ (*[4.4.] Age Reference Operations*).

Linguistic consistencies. Linguistic aspects of the $S_{ub}L_{it}$ topics refer to the specific use of semantics, phonology, and syntax that each individual topic, as well as the topics collectively and consistently, exhibit.

11. Linguistically, consistent conjugations of pronouns are used to connect a literal topic with its $S_{ub}L_{it}$ referent. This is done by tense shifts. For example, the pronouns "this" in the literal statement "/*This*/ 1 Girl Who Was With 2 Guys", instead of lexically selecting "*that*" 1 girl or "*a*" girl who was with two guys" to link the narrative to the actual narrative situation selected the pronoun /*This*/, thus psychologically linking the topic to the actual situation (*[9.1.]. Temporal shift operations*). As to why the older woman was sub-literally referenced as a *girl*, especially since in stories 8 and 9 the literal references to *seniors* and *old* correctly indicated the older age of the woman, based on other protocols, literal terms are often used generically on the $S_{ub}L_{it}$ level. That is, terms like *girl* are used generically as simply a gender reference, not a specific reference to a person's age (*[4.5.] Generic Operations*). The context in which a generic reference is used will typically indicate how the term is meant.

12. Linguistically, consistent conjugations of personal pronouns are used to sub-literally link literal stories to the actual narrative situation. For example, the telling of the literal story /We narrowed them all down to /*3 different options*/ is a $S_{ub}L_{it}$ reference to the *3 dominant members* who narrowed the leadership to themselves. The story was generated by a member of the triad. Methodologically, for members who were not a part of this dominant triad to have introduced this topic with the particular wording using /*We*/ narrowed them down/ would not have been sub-literally congruent with what occurred in the actual narrative situation be-

cause the larger group did not narrow the leadership down to 3 people—the triad did. In addition, for a member of the triad to have said "*they*" narrowed them down, would not have been consistent with their actual situation. This, too, would have meant that the non triad members created the 3 leaders. That topics 1, 7, and 12 were generated by a member of the triad with all others topics being presented by the rest of the younger females is psycho–sociometrically consistent with their affective schemata. These consistencies are refered to as Psycho–sociometric validity (*[4.] Psycho–sociometric Operations*).

13. Nouns are consistently used as adjectives and adverbs to sub-literally link stories to the actual narrative situation. For example, the narrative about, /3 *seniors*/ (literally meaning high school seniors) literally used as a noun, was used as both a noun and an adjective to sub-literally describe the older members of the triadic leadership structure (*[9.2.] Noun Shifts; [4.] Psycho–sociometric Operations).*

14. Prepositions are consistently shifted to their adjectival or adverbial form to sub-literally link stories to the actual narrative situation, as in the statement /being *under* 21/, referring to not being under 21 years old but to being /*under*/ the authority of the triadic leadership structure (*[9.5.] Prepositional Shifts; [5.2.] Dimensional Evaluative Vector Operations; [5.3.] Dimensional Vector Equivalence Operations*).[13]

Associative consistencies. Associative aspects of the $S_{ub}L_{it}$ topics refer to the related aspects of each individual topic as well as the topics collectively being logically and contextually consistent with each other.

15. Most of the narratives containing the number "3" are not just isolated 3's; they are included in or associated with stories about a larger group-like unit, as in rock groups, bars, airplanes, and buses, just as the members of the triad are part of a larger group (*[13.5.] Inclusivity and Exclusivity of Categorical Set Operations*).

16. Semantic associations correspond to the actual *social status* of the triad, as in the phrase /3 *lucky spots*/ found in topic 7 (*[5.1.] Semantic Association Operations*)

17. That the topic /*The 3rd Stream*/ is perhaps a $S_{ub}L_{it}$ reference to the slightly older extremely verbal male is indicated by being associatively congruent with the language community vernacular, meaning someone talking a "steady stream". This is in turn congruent with other literal narratives about people who talk too much (*[5.1.] Semantic Association Operations*; *[9.7.] Vernacular Operations*).

18. In the literal topic /It started snowing *3 hours*/ before the plane arrived at the airport, it needs to be asked again, why both the sub-topic of /*snow*/ and the number /*3*/ were associated? In terms of contextual information adding to the validity of this interpretation is the fact that during the first meeting (the only instructional lecture), the trainer/researcher had *overloaded* them with a sort of crash course on group dynamics. That is, in common vernacular, he "snowed" them with his lecture. So this particular sub-topic about snow is probably a specific $S_{ub}L_{it}$ reference to the trainer/researcher and/or to the fact that the group situation is confusing. Other associated references were a part of this conversation. For example, it was said that the snow caused /*air traffic problems*/ so that they were /*flying around in circles*/ waiting to land. The /air *traffic problems*/ is likely—and typically—a $S_{ub}L_{it}$ reference to communication problems in the literal conversation. For example, not all members had equal "air time", and the discussion rules were not established. The reference to /*flying around in circles*/ has two $S_{ub}L_{it}$ referents. First, it refers to the standard T-group seating arrangement of *sitting a circle*. Second, the conversation was perceived as *not going anywhere*, i.e., in common vernacular, it was just going in circles (*[5.1.] Semantic Association Operations*; *[9.7.] Vernacular Operations*).

19. The topic was about bartenders having the right to not /*serve*/ a person if they have had too many drinks—which was said to be /*for instance, 3 drinks*/—and /*if he sees that you cannot handle more*/. That /*3 drinks*/ were too many was generated by the older woman in the triad and is thus more normative (and predictive from base-rate data on values relative to older age) than a younger person indicating that 3 drinks were too many (*[2.3.] Expectations; [2.4.] Knowledge-base; [4.1.] Resonance Operations; [4.2.] Sociometric Operations below*).[14]

20. The topic /*under 21*/ is dimensionally or spatially consistent with a reference to being /*down behind* Pantry Pride/—a grocery store (*[8.4.] Single number operations*; *[8.5.] Addition operations*). As indicated earlier, on another level of reference, the /being *under* 21/ is a $S_{ub}L_{it}$ reference to the remaining members of the group who were all younger than the 3 leaders. As additional verification, almost immediately connected to this phrase was the statement /Over half of them were *under* age/. The younger female members continuing this topic about people in a bar being under age, were them-

selves underage (*[4.1.] Resonance Operations; [4.2.] Sociometric Operations below*).

21. In the literal topic, /*3 of the 10*/, the individual numbers /*3*/ and /*10*/ correspond to the actual group composition: the /*3*/ leaders and the remaining /*10*/ young female members. As in the first topic, the number /*3*/ gains additional validity by being directly a part of a phrase that included another significant number, the number /*10*/. Combined, the *3 + 10* totals to *13*. The significance of the number 13 is that it corresponds to the exact number of people in the group that day, including the trainer/researcher (*[4.1.] Resonance Operations; [4.2.] Sociometric Operations below, [13.5.] Inclusivity and Exclusivity of Categorical Set Operations*).

22. It is associatively significant that topic 14 about a hypothetical /*John Smith, 21 years old*/ was generated by the younger of the 2 males in the triad (who was about 21 years old), and was not associated as were the other topics involving the prepositions /*under*/, i.e., *under* the control of the triad (*[9.5.] Prepositional Shifts; [5.2.] Dimensional Evaluative Vector Operations; [5.3.] Dimensional Vector Equivalence Operations*).

23. That members not belonging to the triad were subordinate to the triad is associatively reflected in the aspect of topic /*3 weeks ago*/ referring to a large group of people that /were all *down behind*/ a food market called *Pantry Pride* (a statewide supermarket chain). The phrase *Pantry Pride* was associated in a previous session with the older woman in the triad saying she was /*proud*/ i.e., = pride in being a homemaker = Pantry (a small room or closet near a kitchen, in which food, silverware, dishes, etc., are kept). In other words nontriad members were subliterally shown to be subordinate to the older woman. Verification is also indicated by the dimensional vector "/*down*/ behind". The phrase "*they* were all /*down*/ behind", with "they" equaling the larger group membership that was /*under*/ the leadership of the three leaders, one of which was the older woman is important. In terms of verification, invariably, high status, i.e., being a leader, is associated with the dimensional vector /*up*,/ and low status, i.e., being a follower, with /*down*/ (*[5.2.] Dimensional evaluative vector operations; [5.3.] Dimensional vector equivalence operations; [15.4.] Dimensional tracking of deductive subset invariance; [15.5.] Dimensional tracking of permutation invariance*).

24. In the topic /*3 of the 10 People*/, the continuing phrase, /and they *wouldn't serve any of them*/ is a $S_{ub}L_{it}$ reference to the rest of the group not accepting the leadership of the triad, that is, they would not /*serve*/ as followers. This is methodologically important because it is congruent with the remaining topics being associated with negativity toward the triad and that these literal topics were generated by nonmembers of the triad. Conversely topics 1, 7, and 12 were generated by the triad. In short, it would be expected that negative topics would be generated by members who were not a part of the triad (*[4.1.] Resonance Operations; [4.2.] Sociometric Operations below*).

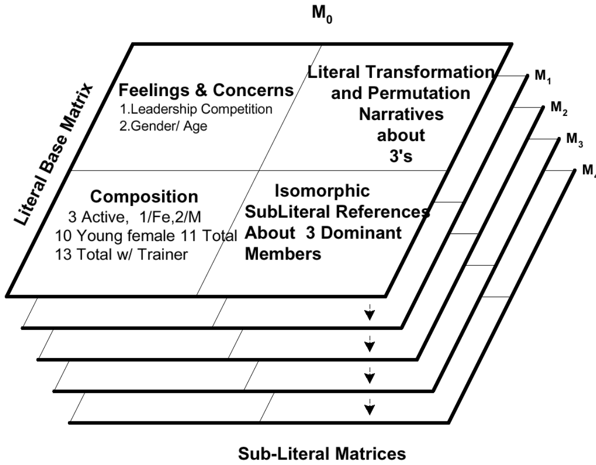

Figure 4. Collapsed lattice matrix.

25. It should be noted that the delineated set of numerically triadic narratives could be expanded to a further set of related numerical references. For example, the fourteenth narrative involving the /*3 of the 10 People*/ is one such non-exclusive reference to the triad, with the number /*10*/ corresponding to the 10 young females in the discussion.

Because the series of numerical references are isomorphic with each other, the matrices can be collapsed into a single compressed cognitive matrix. See Figure 4.

Finally, the significant implication of the above structural, linguistic and associative analysis is that in order for this series of *consistent and structurally integral* numerical topics to occur, each representation and its other consistently associated aspects(i.e., age, gender) and corresponding isomorphic meanings across the various transformations and permutations must somehow be cognitively (a) mapped, (b) tracked, and (c) stacked systemically throughout multiple levels of meaning and through the various story permutations, all remaining invariant with respect to the specific set of characteristics (e.g., numerical magnitude, age, gender, etc.) and meanings, *[14.] Matrix, and Lattice Structure Validation Operations; [15.] Multicorrelative Transformational Validation*).

A Logico–Formal and Idealized Systemic Summary of Logico–Mathematic Operations and Validation Structure

Methodologically, validation is based on (1) internal relationships of invariance constituting formally consistent sets and transformations of cognitive structures similar in form to establishing arithmetic proofs and (2) external relationships of isomorphic correspondences to empirical group interaction data. Thus, the method is neither causal nor linear but based on internal and external consistency and the invariance of a nomological formal structure (see below). There are two essential levels of validation, the internal and external. Similar to mathematical structure, internally they are integral, systemic, and consistent isomorphic order structures. These yield logical coherence and reliability. Second, are mappings and matchings of external contexts to these internal corresponding structures in the same way that the exponential formula corresponds to the growth of populations, and a host of other real things in the world.

Though not an exhaustive list, the following summary has been abstracted from the complete array of cognitive operations and validation structures that constitute the LMS methodology (HASKELL 2003a). It is an idealized summary of selected operations that demonstrate the systemic, integral, correspondent, and logically consistent structure of the operations and structural relations.

1. The general meaning of a literal topic X is found to *match* the concrete narrative situation (*[1.1.] Matching Operations*).

2. A finer grain analysis also shows that topic X with its further specific characteristics and attributes $_{a\,b\,c\ldots n}$, are found to *map* onto, and be isomorphic with, the actual narrative situation (*[1.2.] Isomorphic Mapping Operations*).

3. In addition, the matching and mapping are found to be congruent and consistent with what is contextually, historically, and developmentally known about (a) the narrative situation and, (b) that topic X analogically corresponds to the known expectations and affective schemata of the particular members generating the narratives (*[2.1.] Historical, [2.2.] Developmental Stage; [2.3.] Expectations*).

4. Further, the content of topic, $X_{a\,b\,c\ldots n}$, is congruent and consistent with the physical and psychological selection constraints of the narrative situation (*[2.6.] Selection Response Field*), where out of all possible lexical and physical selection alternatives, the specific lexical and physical selection alternatives are consistent with the topic.

5. Continued matching and mapping operations demonstrate additional sets of literal topics $X'_{a\,b\,c\ldots n}$ and $X''_{a\,b\,c\ldots n}$, found to be transformations of the original topic X, with each expressing the same general affective schema but using different topic content or surface–structure representations (*[3.1.] Transformational Operations*). These are *transformations* of the literal topic.

6. Other topics are found to be *permutations* of X, i.e., $X^1_{a\,b\,c\ldots n}$, $X^2_{a\,b\,c\ldots n}$, with each expressing a different aspect of the same basic affective schema (*[3.2.] Permutational Operations*). In support of this finding, the concern expressed in each permutation topic is found to belong to the same logical class and corresponds to the affective schema in the literal narrative situation (*[3.3.] Transitional Narratives)*, as indicated by specific linguistic expressions (*[3.4.] Transitional Linkages*).

7. Consistently, discussants generating literal topic X are those discussants experiencing the affective schema about the actual narrative situation that the $S_{ub}L_{it}$ referent or topic X expresses (*[4.1.] Resonance Operations*).

8. Conversely to the above, discussants generating the transformational topics $X'_{a\,b\,c\ldots n}$ and $X''_{a\,b\,c\ldots n}$ containing negative $S_{ub}L_{it}$ attributions are not a part of the literal narrative situation to which the negative attribution is a $S_{ub}L_{it}$ reference (*[4.2.] Sociometric Operations*).

9. Similarly, discussants generating literal topics containing *gender* and/or *age* references and distinctions are the discussants who experienced gender and age affective schemata in the literal narrative situation expressed by the hypothesized $S_{ub}L_{it}$ referent of the topic (*[4.3.] Gender Reference Operations*).

10. Significantly, and in summary up to this point, based on affective and motivational considerations, all sets of narratives are psycho–dynamically, sociometrically (*[4.] Psycho–sociometric Operations*) and isomorphically correspondent and logically consistent. This is indicated by: If $S_{ub}L_{it}$ referents are generated from affective schemata, it follows that discussants generating literal topics with hypothesized $S_{ub}L_{it}$ referents are those experiencing the schema in the actual narrative situation that the $S_{ub}L_{it}$ meaning of the topic expresses.

11. Other semantic operations (*[5.1.] Semantic Association Operations*), to topic X are also found to be logically and structurally congruent with all of the above operations. For example, dimensional (prepositional) vectors associated with each topic transformation and its permutations are consistent

with the literal narrative vectors (*[5.2.] Dimensional Evaluative Vector Operations; [5.3.] Dimensional Vector Equivalence Operations*), i.e., if the content of literal topic *X* is associated with being "up", meaning high status, then all of the transformations, i.e., $X'_{a\,b\,c...n}$ and $X''_{a\,b\,c...n}$, and permutation, i.e., $X^1_{a\,b\,c...n}$, $X^2_{a\,b\,c...n}$ narratives, should also be associated with the vector "*up*". This is especially significant for validation.

12. In addition, unconsciously generated physical gestures that a discussant exhibits during the relating of a literal narrative may correspond to the $S_{ub}L_{it}$ referent of the literal topic (*[6.1.] Ocular Operations; [6.2.] Gestural Operations*). For example, automatic activated minute ocular and/or hand gestures may be directed toward the narrative member about which the topic is $S_{ub}L_{it}$.

13. Names and initials in literal topic *X* and their transformations and permutations also match and map onto discussants names and initials in the narrative situation (*[7.1.1.] Initials; [7.1.2.] Embedding; [7.1.3.] Fusions*) as indicated by a number of consistent and corresponding semantic and phonological operations, e.g., *[8.1.1.] Portmanteaus; [8.1.2.] Homophonic Operation; [8.1.3.] Oronymic Operations; [8.1.4.] Paronymic Operations,* as well as syntactic *[8.2.] Syntactic Ordering Operations*, along with shifting and tagging operations *[9.] Linguistic Shifting, and Tagging Operations; [9.1.] Temporal Shift Operations; [9.2.] Noun Shifts; [9.3.] Adjectival Shifts; [9.4.] Plural/Singular Shifts; [9.5.] Prepositional Shifts; [9.6.] Tagging Operations*.

14. Consistent within and among topic transformations and permutations, and consistent with other operations as indicated above, $S_{ub}L_{it}$ referents are generated by consistent reversal and inversion operations (*[10.] Reversal, Inversion, Opposition Operations*), as indicated by textual content (*[10.1.1.] Textual Expressions*), as well as strategic memory distortions that are consistent with the above operations, (*[11.] Memorial and Perceptual Operations*).

15. Just as with semantically expressed topics, numerical expressions within topic *X*, and its transformations and permutations are found to be a consistent set of $S_{ub}L_{it}$ representations of the membership sub-grouping factions within the narrative situation (*[12.] Arithmetical Operations*). This is partly indicated by the numerical values isomorphically mapping on to and thus matching the narrative numerical membership composition, and as supported by the particular numerical values expressed being consistently re computed in subsequent topics to match changes in membership or sub-grouping occurrences (*[12.1.] Numerical Recomputation*], as well as by the numerical representations consistently matching the narrative subgroup factions by gender and other relevant demographics (*[13.] Logico–mathematic Operations*).

16. As a further validation process, given the consistency and internal logical structure of the entire array of linguistic and cognitive operations, it follows that with any given narrative analysis its isomorphic structural and $S_{ub}L_{it}$ characteristics should be retrodictive to numeric narratives in previous sessions and predictive of numeric narratives in future sessions (*[18.] Retrodiction and Prediction*).

17. In addition, the cognitive, linguistic and logical operations found in topic *X* and its transformations and permutations are found to be individually and logically consistent with each other, and to additional higher order structural isomorphic inter systemic mappings. This internal systemic consistency provides structural validation. In analyzing and validating the internal cognitive structure manifested in a set of narratives and their transformations and permutations as well as their various affective aspects from which the $S_{ub}L_{it}$ levels are partially derived, each topic can be assigned to cells within an isomorphic cognitive matrix series (*[14.1]. Matrix Structures*). Together these matrices form a kind of cognitive lattice structure composed of the base matrix (M_0), and a series of transformational matrices (M_1, M_2, M_3, M_4) the cells of which contain the different levels of $S_{ub}L_{it}$ material that isomorphically correspond to the cells in the generative base or literal matrix. Each tier can be seen as a harmonic of the others (*[14.2.] Lattice Structures*). See previous figures.

18. Finally, validation of the topic X series involves systemic intra- and inter-narrative multicorrelative transforms (*[15.] Multicorrelative Transformational Validation*), the basics operations of which involve logically deductive sets such that if a literal narrative *X* is consistently associated with a given characteristic or set of characteristics, i.e., $_{a\,b\,c...n}$, (*[5.] Associational and Dimensional Operations*), then it follows that the corresponding $S_{ub}L_{it}$ subsets of transformations, or permutations of narrative *X*, i.e., $X'_{a\,b\,c...n}$, $X''_{a\,b\,c...n}$ and its permutations, $X^1_{a\,b\,c...n}$, $X^2_{a\,b\,c...n}$ should consistently exhibit the identical characteristics and other associated attributes which can thus be tracked throughout the entire series of lattice-like matrices.[15] Again, if these consistent and integral logico–mathematic-structural operations are valid, then the (1) mapping, (2) tracking, and (3) isomorphic

stacking structures involved in these operations demand explanation.

Any contraindication within this idealized analysis would indicate a degree of methodological invalidation of the particular operation (see, *[16.] Nomological Net work*).[16] Finally, given the correlationally integral, systemic, transformational, and permutational numeric structures found with the above series of numerical narratives, along with the finding that this series conforms to an algebraic structure (HASKELL/BADALAMENTI in press) it seem reasonable to suggest that these *numerical* finding lend their validational force to the less concise *semantic* analysis of $S_{ub}L_{it}$. narratives. Cognitively, it is reasonable to assume that the structure of semantic narratives would be subject to the same undergirding structure of numerical narratives (see HASKELL 2003a, *First Extended Exemplification: Phonetic and Syntactic Structure, [8.1.3.] Phonetic Operations; [8.2.] Syntactic Ordering Operations* for a highly structured semantic instantiation).

Extraordinary Claims and Methodological Design

Many of the cognitive operations and $S_{ub}L_{it}$ findings seemingly make extraordinary claims. Accordingly, it is a principle in science that extraordinary instantiation and corroboration for such claims is required. Agreeing with this principle, the LMS method was developed to control and provide validation procedures for $S_{ub}L_{it}$ findings. However, the method and its operations also tend to be seen as making extraordinary claims. While this paper can not provide extensive corroborations the following summary of arguments and supporting findings presented in more detail elsewhere (HASKELL 1991, 2003a) will provide initial and illustrative support for the LMS operations and numeric $S_{ub}L_{it}$ claims.

Two major related issues that have been raised about the LMS method and its $S_{ub}L_{it}$ findings are that conventional experimental methodology and sampling procedures are not utilized, and thus may be perceived as an absence of sound controlling procedures. One reason for these critiques is that the LMS method and its $S_{ub}L_{it}$ findings engender a kind of epistemological paradigm shift in criteria for what constitutes a valid controlling methodology.[17]

Suffice it here to say that two models related to the LMS methodology require neither experimental design nor sampling procedures. These models are mathematics and structural linguistics. The problem here is similar to that dealt with in the study of linguistics. Linguistics did not develop its theories of syntax using experimental design and chi-square tests. It is based a structural, inferential, and lawful system of relations.

The use of a contingency table and a chi-square test is only appropriate when attempting to show (or refute) that two factors are related. Statistical analysis is typically not appropriate for lawful mechanisms. Its proper use is in estimating parameters or in testing for the probable presence or absence of relationships when the observables are regarded as random or at least as highly variant. Sampling and other statistical methods, then, do not apply where regularity or lawfulness is presumed to exist. In dealing with linguistic and with verbal language (speech), lawfulness is assumed. For example, the concept of speech parts and their combinations such as nouns, verbs, adverbs and so on is regarded as neither random nor merely probable but as lawful to language. Likewise, the concept of the *meaning* of a sentence constructed with the speech parts is not regarded as random or probable (from the point of view of the speaker) even though the sentence construction itself may show some variation.[18] The methodological point here is that experimental/statistical methods are neither necessary nor sufficient for investigating the complex linguistic processes briefly outlined here just as they are not relevant in linguistics or in mathematics where inferential and logical procedures are used for analysis and validation.

Background Knowledge

An extraordinary claim is only extraordinary to the extent it does not seem to cohere with an accepted paradigm. It is suggested here that, in part, that a lack of background knowledge is also responsible for what is considered extraordinary about the LMS method and its findings (see HASKELL 2003b, 2003c). Unlike with novel methods, with experimental and statistical methods, the background knowledge and assumptions undergirding them are not an issue; they function as consensual "givens" within a community of researchers. The concept of an "extraordinary claim", then, is one relative to a given knowledge-base.

In the absence of an appropriate and pertinent knowledge-base, then, some of the seemingly anomalous methodological operations are perceived as not having face validity and therefore credibility. Among the primary background knowledge required to accept the credibility of the logico–mathematic method and its findings is a knowledge of structuralist epistemology (see HASKELL 2003c). Other areas of background knowledge include the vast research on metaphorical and analogical reasoning and mapping processes (e.g., GENTNER 1983; HASKELL 1987a, 1987b; LAKOFF/JOHNSON 1980; MACCORMAC 1985; VOSNIADOU/ORTONY 1989). Narratives and their $S_{ub}L_{it}$ referents can be conceptualized as unconscious metaphorical or analogical reasoning and mapping processes about extant affective schemata. analogically notated the numeric series presented above look as follows. *The series of literal topics about 3's : their transformations and permutations :: the 3 dominant members in the discission : their various characteristics.*

Theoretical Bases

Another related major issue about the LMS method and its $S_{ub}L_{it}$ findings involves their not having a clear theoretical or explanatory base. While the integral set of cognitive and psycho–linguistic operations constituting the LMS methodology stands on its own, various parts of a theoretical base can be found in HASKELL (2003c, 2000, 1989). This theoretical base involves a neurological transformation of invariance function subserving not only $S_{ub}L_{it}$ but other phenomena such as analogical reasoning and transfer, similarity relations, etc. (2000).[19] Lacking a clearly delineated theory, a surface reading of the LMS method, and given the seemingly anomalous findings generated by motivated unconscious "psychodynamics", $S_{ub}L_{it}$ findings are often perceived as somehow based on a "closet" variant of psychoanalytic theory. Again, that the LMS method stands independently on its own—*qua method*—notwithstanding, a brief outline of some research that can be seen as theoretical bases are in order.

Masked Priming

Research on masked priming and the automatic activation of unconscious schemas can provide an initial and partial theoretical base (see HASKELL 2003c) for $S_{ub}L_{it}$ findings. Masked priming involves stimuli presented either outside of a subject's focal attention, or are presented so rapidly that they can not be consciously perceived but nevertheless influence subsequent responses to stimuli.[20] The research on automatic versus controlled activation of schemata indicates that stimuli presented during an experiment—or stimuli present in the everyday environment—can elicit automatic responses. This automation is analogous to the automatic functioning of highly practiced motor skills. Based on the automatic versus controlled processing research, BARGH/BARNDOLLAR (1996) suggest that under certain conditions unconscious goals and motives can be automatically activated. They view the unconscious as the repository of goals and motives (schemata) that with frequent use or repetition become automated strategies for responding to the environment. These goals and motives can be activated by environmental situations that function as priming stimuli. BARGH has termed this process the "auto-motive" model and suggests that much of conscious behavior is the consequence of the automatic activation of chronic goals and motives (see BARGH/CHARTRAND 1999, for synopsis). Masked priming techniques have been utilized to automatically activate racial stereotypes (e.g., DEVINE 1989; WITTENBRINK/JUDD/PARK 1997) and other schema-based material such as social affiliation and achievement behaviors (e.g., BARGH/GOLLWITZER 1994). In priming and automatic activation terms, then, events within narrative situations can function as priming stimuli that automatically activate affective schemata about those situations resulting in $S_{ub}L_{it}$ material.

Regarding the seemingly controversial claim that numbers contained in literal narratives have $S_{ub}L_{it}$ referents (HASKELL 1983, 2003a), along with the above research as a model the following can be seen as providing a theoretical base. Findings show that animals (BOYSEN/BERNSTSON 1996; CAPALDI/MILLER 1988; CHURCH/MECK 1984; KOEHLER 1951; MOYER/LANDAUER 1967; WOODRUFF/PREMACK 1981) engage in numerical mapping, tracking, and calculation of up to four objects. Further, WYNN's, pioneering work (WYNN 1992; CHIANG/WYNN 2000) demonstrating that months-old infants engage in numerical mapping, tracking, and calculation of up to four objects (see also DEHAENE 1997; KOECHLIN/DEHAENE/MEHLER 1997). Certainly this animal "numerosity", as it is called is not a "conscious" process.

In addition, neurological research on unconscious numerical processing provides support as does the work of THOMPSON et al. (1970). There seem to be specifically dedicated neurons and entire neurological circuits for various numerical operations like addition, subtraction, and approximation (DEHAENE/CHANGEUX 1993; STANESCU-COSSON et al. 2000). Using event-related potentials (ERPs) and functional magnetic resonance imaging (fMRI) measures, along with a masked priming methodology it has been demonstrated that consciously unperceived numerical stimuli not only activate visual circuits associated with numbers but that unconscious processing is being carried out as well (DEHAENE et al. 1998; DEHAENE et al. 2001; NACCACHE/DEHAENE 2001). At about *190* milliseconds numerical quantity is encoded and processed, indicating that unconscious numerical processing is occurring. Given that numerical calculations are found to have specific neurological pathways, and that numeric processing can occur unconsciously, it lends support for the veridicality of unconscious numeric mapping, calculation, and approximations found in verbal narratives. In other words, that unconscious numerical cognition is not just "psychological" and demonstrated only by behavioral data, but has a basis in corresponding neurological substrates.

Finally as noted above, HASKELL/BADALAMENTI (in press) have recently found that the series of $S_{ub}L_{it}$ numeric topics analyzed here exhibits an algebraic structure, suggesting an algebraic structure to cognition. Further, based on his analysis of connectionism and computational data MARCUS (2001) suggests the brain may inherently function algebraically. Relatedly, DEHAENE notes the work of HITTMAIR-DELAZER/SEMENZA/DENES (1994) and HITTMAIR-DELAZER/SAILER/BENKE (1995) who found that when people become acalculic they do not necessarily lose their knowledge of algebra. Though there are few such findings, they suggest that there exist neuronal circuits responsible for algebraic functions that are largely independent of those involved in mental calculation

The Evolutionary Social Psychology of $S_{ub}L_{it}$ Cognition

Given the animal and neurological findings for unconscious numerical cognition, it would be reasonable to assume an evolutionary base. From an evolutionary biological perspective, the development of unconscious cognitive skills would have had important survival benefits for hominids. Various evolutionary researchers (DUNBAR 1996; CORDS 1997) have argued the importance of animal and early human ability for tracking group membership, for assessing alliances, disputes, absences, and other social dynamics.

Just as an evolutionary analysis would predict, virtually all unconscious mapping and calculating found in the $S_{ub}L_{it}$ stories and topics were about tracking group membership both in terms of the total number of members as well as for dominance relations, gender, sexual tensions, age differentials among members, and for tracking favors, all equally important variables for animals functioning in a social environment of increasing complexity. It has been known for some time that animals from vampire bats (WILKINSON 1984) to primates (SEYFARTH/CHENEY 1984) engage in reciprocal favor giving and tracking (see BUSS 1999).

A question that arises is: Why would numerical tracking be carried out unconsciously rather than consciously? One answer to this question is that basic numerical tracking skills likely evolved prior to the advent of modern humans, and thus laid neurological circuits independent of "conscious" thought and language. A more serious question is: Why would numerical tracking and communication about feelings and concerns be carried out unconsciously after modern human consciousness and language evolved?[21] The answer may involve the strategic deception of others, perhaps analogous to animals evolving camouflaging and other deceptive practices to adapt to the interpersonal political realities of social living.

TRIVERS (1971) well-known research on reciprocal altruism showing that non human animals must devise strategies for detecting "cheating" on payments due for past favors, suggests that animals must devise strategies for hiding their cheating. The argument is that consciously trying to hide the fact of cheating, is not all that effective. Little cues are leaked out to others. The same strategy is evident in humans. In other words—and contrary to what is generally thought—humans are not all that effective at consciously lying. So to resolve this problem, humans must deceive themselves, so that consciously they believe what they are saying. According, self deception allows for a more effective deceiving of others.

Conclusion

Certainly, the evolutionary numeric precursors are not identical to their evolved higher-order $S_{ub}L_{it}$ incarnations. The numerical findings in $S_{ub}L_{it}$ verbal narratives are more sophisticated than their presumably hard-wired precursors. How might that increased sophistication be explained? First, $S_{ub}L_{it}$ numerical referents, unlike the numeric tracking of animals and infants, were generated by adults who had been exposed to learning some mathematical operations and who had more experience in encoding spatio–temporal situations. As a consequence adults have developed a more integral set of related neurological circuitry. More importantly, they are more sophisticated than their precursors because both unconscious and conscious cognitive processes are likely involved in generating them. As BARGH/BARNDOLLAR (1996) have insightfully observed, a widespread assumption in the research on unconscious processing has been that conscious and unconscious processes are distinct, or at least that the lower level processes exert their influence serially not simultaneously.

Considerable research, including $S_{ub}L_{it}$ findings, suggests that both conscious and unconscious processes influence each other with reciprocal reentry feedback loops. This latter model is more congruent with a biological and evolutionary view of brain functioning (EDELMAN 1992), and with what KOSSLYN/KOENIG (1995) have termed a "wet brain" approach to cognition as opposed to "dry" computational models.[22] Thus, sophisticated numerical mapping, tracking, and calculation found in $S_{ub}L_{it}$ narratives make sense if the mapping, tracking, and calculation are generated by multiple neurological circuits on both conscious and unconscious levels operating simultaneously. One model for this processing is the multiple levels of unconscious circuitry involved in the conscious experience of vision laid out by MARR (1983). Another partial model is TRIESMAN's (1985) preattentive processing.

Although some $S_{ub}L_{it}$ narratives and operations strain common sense and standard understanding of language and mental processes, so do many other anomalous phenomena that we yet do not understand. Certainly the procedural and cognitive operations presented here are neither typical nor easily amenable to standard cognitive or linguistic methodology. The robustness of the findings however warrant further programmatic research. Continued methodological and theoretical research needs to be developed to expand our understanding of $S_{ub}L_{it}$ phenomena and their relationships to areas already developed in psycholinguistics and cognitive science.

Author's address

Robert E. Haskell, New England Institute of Cognitive Science and Evolutionary Psychology, University of New England, Biddeford, Maine 04005, USA
Email: haskellr@maine.rr.com

Notes

1 The methodology presented here, along with its theoretical bases have been in development for some time (HASKELL 1978, 1982, 1983, 1984, 1985, 1986, 1987a, 1987b, 1987c, 1988, 1990–1991, 1991, 1989, 1999a, 1999b, 2000, 2002, 2003d).

2 Because of the seemingly anomalous character of LMS methodology and its findings, there are a host of issues that need to be understood but which can not be addressed here. One of these issues is the historical apparent distinction between literal versus figurative language. In linguistics the traditional distinction between "literal" and "figurative" meaning has been shown to be problematic if not untenable (see ARIEL 2002; HASKELL 2002). For purposes of this paper the term "literal" should be understood as "conventional". Similarly, when linguistic labels and notions are noted here, e.g., *[9.2.] Noun Shift*, they are to be understood as just conventional linguistic notions used for explanatory purposes, not as real "shifts" that occur in cognitive and conceptual structures; it is unlikely that the cognitive/brain apparatuses generating language recognize such linguistic explanatory models.

3 Historically, the analysis of unconscious or unintended meaning has its intuitive origins in everyday language with puns and slips-of-the-tongue; its initial conceptual roots, however, lay in a sporadic array of primitive precursors from clinical psychodynamics (e.g., FREUD 1960), and as metaphorical-like language within a psychoanalytic framework. Derived from these clinical contexts, various historical approaches in small group dynamics research has conceptualized unconscious meaning in narratives as "fantasy theme analysis" in the classic work of the social psychologist BALES (1970) and his associates issuing out of the classic Harvard Social Relations 120 small group laboratory course during the 1960s.

4 For example, the long history of "metaphor" and "analogy" (processes) until the past few decades were viewed as simply "literary" devices by cognitive science. Metaphorical and analogical reasoning are now recognized as being central to understanding cognition (GENTNER 1983; HASKELL 1987a, 1987b; LAKOFF/JOHNSON 1980; MACCORMAC 1985; VOSNIADOU/ORTONY 1989). Just as "metaphor" is a problem for traditional linguistic theory (–ies, except as idioms), $S_{ub}L_{it}$ findings are even more problematic. As I was completing this paper, however, I became aware of JACKENDOFF's (2002) seminal reconstruction of linguistic theory. Even more than Langacker's cognitive grammar approach to linguistic processes, $S_{ub}L_{it}$ findings seem to be more compatible and consistent with JACKENDOFF's framework. He not only bridges the traditional "modular" barriers between syntax, phonology, and semantics but he connects a "theoretical linguistics" with everyday language by incorporating the hoary problems of "meaning" and context, all within a clearly psychological and neuroscience perspective; JACKENDOFF's reconstruction of linguistic processes render some of the more seemingly anomalous linguistic operations demonstrated in this methodology now theoretically more explainable (see in particular, pp202–230).

5 The concept of a cognitive unconscious has been developed in recent years to distance rigorous research from various conceptualizations of a FREUDIAN unconscious (see, KIHLSTROM 1984; KIHLSTROM et al. 2000; PIAGET 1973; REBER 1993). The notion of unconscious processes existed, of course, in both philosophy and psychology long before FREUD (see ELLENBERGER 1970; SMITH 1999, 1991; WHYTE 1978). As used by HASKELL, the concept is conceptualized as both a "cognitive and emotional unconscious". A more precise definition, notwithstanding, the new cognitive unconscious includes the system of operations found in HASKELL (1991, 2003a). For ease of exposition, however, henceforth the term "unconscious" should be understood in this sense.

6 A "cognitive operation" is defined as an internalized reversible action carried out in thought and which is part of a larger and integral set of cognitive structures and processes. Arithmetically, addition is an example. Subtraction is the reverse of addition; it is the same operation carried out in the opposite direction. One can be added to 1 and get 2; then 1 can be subtracted from 2 to return to 1 gain. In more phenomenological terms, a Gedankenexperiment is a cognitive operation: one can imagine or conceive of an experiment and manipulate it in ones head and reverse its order.

7 The references in brackets, e.g., (*[12.] Arithmetical*) refer to sections in the complete methodology (HASKELL 1991, 2003a).

8 A transcribed protocol of this eighty-minute session is available for research purposes upon request.

9 Only rarely does the $S_{ub}L_{it}$ referent of the literal topic evolve to a semi conscious level. When it does occur, it seems to be associated with linguistic shifts (*[9.1.] Shifting Operations*).

10 Because it is visually easier to follow the numerical analysis if numbers are presented as numerals, e.g., "3", they will not be represented in typical orthographic style as the word "three".

11 Methodologically, it is significant to note that I have found the repeated occurrences, say, of the number 5 in literal stories in a conversation where only 5 people are active will change to 4 if one of the members leaves or is otherwise absent. Moreover, I have found that these numbers in literal stories and their changes are consistent across different sessions of the same discussion group.

12 It should be pointed out that it is possible these transformations are really permutations. This would not, however, change the topics exhibiting the set or mathematical properties; it would only change the content of the sets or groups.

13 Prepositional phrases are composed of a preposition and its object, which often have adjectival or adverbial properties (*[9.5.] Prepositional Shifts*).

14 Previous versions of this particular narrative were partially incorrect. In returning to the protocol, the member generating this narrative was the older woman, not a younger member of the group. This correction, however, does not change the essential $S_{ub}L_{it}$ referent of the narrative, only its validation properties.

15 While the linguistic and cognitive structures of $S_{ub}L_{it}$ narratives are consistent and determined by rules, the rules are often governed by the context in which they are used. If occasionally there seem to be inconsistencies and contradictions, it is likely because all of the context-dependent rules have yet to be discovered.

16 See *[16.] Nomological Validation Network [17.] Falsification [18.] Retrodiction and Prediction*. Derived from the Greek term "nomos" meaning "law". More recently the concept refers to the discovery of laws within a network of logical relations. Not all $S_{ub}L_{it}$ referents can be integrally validated with the complete array of operations. Validation and falsification procedures, then, derive from being sufficiently tied to an integral network of procedures and operations that exhibit a high degree of formal adequacy of the correspondences and the plausibility of the entire analysis. Thus, logico–mathematic-structural validation is a network concept comprised of varying degrees of interconnecting relationships. Each hypothesized $S_{ub}L_{it}$ narrative is thus evaluated, not in isolation, but in relation to direct and indirect relationships and procedures in the network of logically consistent and systemic cognitive and linguistic operations. A nomological network is a form of construct validity. FEIGL (1956, after CRONBACH/MEEHL 1955) describes such a system as a *nomological network*.

17 I wish to thank David L. SMITH, a philosophy colleague for his conceptualizing the series of issues I had been addressing as constituting a paradigm shift.

18 I would like to thank Tony BADALAMENTI for further clarifying this methodological issue in relation to my methodology.

19 See Chapter Eleven.

20 There has been debate whether masked priming can activate cognitive processes without first being accessed through consciousness. Using event related potentials (ERPs) and functional magnetic resonance imaging (fMRI), NACCACHE/DEHAENE (2001a) and DEHAENE et al. (1998) have recently demonstrated that unconsciously received stimuli can activate unconscious cognitive processing.

21 I wish to thank my colleague David L. SMITH for leading me to the evolutionary literature for the possible origins of $S_{ub}L_{it}$ phenomena. He is currently in the process of developing an evolutionary framework for $S_{ub}L_{it}$ communication.

22 While it is acceptable for certain research purposes to ignore context and meaning and thus engage in what FODOR (1980) has labeled (apparently after CARNAP) "Methodological Solipsism", or to act "As-If" (VAIHINGER 1924) the world was somehow equivalent to ones theories, it is another to confuse this pragmatic research strategy with "reality". In addition to GARDNER's (1985) early critique, it is increasing recognized (EDELMAN 1992; KOSSLYN/KOENIG 1995) that computationalism(s)—despite its success in certain areas for certain purposes—are not how the mind works. Computational approaches to cognitive processes —with their increasing number of epicycles—seem at best to be ingenious PTOLEMAIC models of how the mind works.

References

Ariel, M. (2002) The demise of a unique concept of literal meaning. Journal of Pragmatics 34(4):361–402.

Baars, B. J./Cohen, J./Bower, G. H./Berry, J. W. (1992) Some caveats on testing the Freudian slip hypothesis: Problems in systematic replication. In: Baars, B. J. (ed) Experimental slips and human error. Plenum Press: New York, pp. 289–313.

Bales, R. F. (1970) Personality and interpersonal behavior. Holt, Rinehart and Winston: New York.

Bargh, J. A./Barndollar, K. (1995) Automaticity in action: The unconscious as repository of chronic goals and motives. In: Gollwitzer, P. M./Bargh, J. A. (eds) The psychology of action. Guilford: New York, pp. 457–471.

Bargh, J. A./Chartrand, J. L. (1999) The unbearable automaticity of Being. American Psychologist 54:462–479

Bargh, J. A./Gollwitzer, P. M. (1994) Environmental control of goal-directed action: Automatic and strategic contingencies between situations and behavior. In: Spaulding, W. D. (ed) Nebraska symposium on motivation. Vol. 41: Integrative views of motivation, cognition, and emotion. University of Nebraska Press: Lincoln, pp. 71–124.

Boysen, S. T./Bernstson, G. G. (1996) Quantity-based interference and symbolic representation in chimpanzees (Pan trogolodytes). Journal of Experimental Psychology: Animal Behavior Processes 22:76–86.

Buss, D. M. (1999) Evolutionary Psychology: The new science of the mind. Allyn and Bacon: Boston.

Capaldi, E. J./Miller, D. J. (1988) Counting in rats: Its functional significance and the independent cognitive processes that constitute it. Journal of Experimental Psychology: Animal Behavior Processes 14:3–17.

Cronbach, L./Meehl, P. (1955) Construct validity in psychological tests. Psychological Bulletin 52(4):281–302.

Chiang, W./Wynn, K. (2000) Infants' tracking of objects and collections. Cognition 77(3):169–195.

Church, R. M./Meck, W. H. (1984) The numerical attribute of stimuli. In: Roitblatt, H. L./Bever, T. G./Terrace, H. S. (eds) Animal cognition. Erlbaum: Hillsdale NJ, pp. 445–464.

Cords, M. (1997) Friendships, reciprocity, alliances, and repair. In: Whiten, A./Byrne, R. W. (eds) Machiavellian intelligence II: Extensions and evaluations. Cambridge University Press: New York, pp. 24–49.

Dehaene, S. (1997) The number sense: How the mind creates mathematics. Oxford University Press: New York.

Dehaene, S./Changeux, J.-P. (1993) Development of elementary numerical abilities: A neuronal model. Journal of Cognitive Neuroscience 5(4):390–407.

Dehaene, S./Naccache, L./Cohen, L./Le Bihan, D./Mangin, J. F./Poline, J. B./Riviere, D. (2001) Cerebral mechanisms of word masking and unconscious repetition priming. Nature Neuroscience Special Issue 4(7):752–758.

Dehaene, S./Naccache, L./Le Clec'H, G./Koechlin, E./Mueller, M./Dehaene, L. G./van de Moortele, P. F./Le Bihan, D. (1998) Imaging unconscious semantic priming. Nature 395(6702):597–600.

Devine, P. G. (1989) Prejudice and out group perception. In: Tesser, A. (ed) Advanced social psychology. McGraw Hill: New York, pp. 468–524.

Dunbar, R. (1996) Grooming, gossip, and the evolution of language. Harvard University Press: Cambridge MA.

Edelman, G. (1992) Bright air, brilliant fire: On the matter of mind. Basic Books: New York.

Ellenberger, H. (1970) The discovery of the unconscious: The history and evolution of dynamic psychiatry. Basic Books: New York.

Feigl, H. (1956) Some major issues and developments in philosophy of science of logical empiricism. In: Feigl, H./Scriven, M. (eds) Minnesota studies in the philosophy of science. University of Minnesota Press: Minneapolis, pp. 3–37.

Fodor, J. A. (1980) Methodological solipsism considered as a research strategy in cognitive psychology. Behavioral and Brain Sciences 3:63–109.

Freud, S. (1960) The psychopathology of everyday life. The standard edition of the complete psychological works of Sigmund Freud (Translated by J. Strachey), Volume 6. W. W. Norton: New York. Originally published in 1901.

Fromkin, V. A. (1973) Speech errors as linguistic evidence. Mouton: The Hague.

Gardner, H. (1985) The mind's new science: A history of the cognitive revolution. Basic Books: New York.

Gentner, D. (1983) Structure mapping: A theoretical framework for Analogy. Cognitive Science 7:155–170.

Haskell, R. E. (1978) An analogic model of small group behavior. International Journal of Group Psychotherapy 28:27–54.

Haskell, R. E. (1982) The matrix of group talk: An empirical method of analysis and validation. Small Group Behavior 13:165–191.

Haskell, R. E. (1983) Cognitive structure and transformation: An empirical model of the psycholinguistic function of numbers in discourse. Small Group Behavior 14:419–443.

Haskell, R. E. (1984) Empirical structures of mind: Cognition, linguistics, and transformation. The Journal of Mind and Behavior 5:29–48.

Haskell, R. E. (1985) Thought-things: Levi-Strauss and the modem mind. Semiotica 55:1–17

Haskell, R. E. (1986) Social cognition and the non-conscious expression of racial ideology. Imagination, Cognition and Personality 6:75–97.

Haskell, R. E. (ed) (1987a) Cognition and symbolic structures. Ablex: Norwood NJ.

Haskell, R. E. (1987b) Cognitive psychology and the problem of symbolic cognition. In: Haskell, R. E. (ed) Cognition and symbolic structures. Ablex: Norwood NJ, pp. 85–102.

Haskell, R. E. (1987c) Structural metaphor and cognition. In: Haskell, R. E. (ed) Cognition and symbolic structures: The Psychology of metaphoric Transformation. Ablex: Norwood NJ, pp. 241–255.

Haskell, R. E. (1988) Small group "fantasy theme" analysis: Anthropology and psychology, a comparative study of the psychosocial structure of a ritual ceremony. Journal of Psychohistory 16:61–78.

Haskell, R. E. (1989) Analogical transforms: A cognitive theory of the origin and development of equivalence transformation, Part I, II. LMS and Symbolic Activity 4:247–277.

Haskell, R. E. (1990–1991) Cognitive operations and nonconscious processing in dreams and waking verbal reports. Imagination, Cognition and Personality 10:65–84.

Haskell, R. E. (1991) An analogical methodology for analysis and validation of anomalous cognitive and linguistic operations in small group (fantasy theme) reports. Small Group Research 22:443–474.

Haskell, R. E. (1999a) Between the lines: Unconscious meaning in everyday conversation. Plenum/Insight Books: New York.

Haskell, R. E. (1999b) Unconscious communication: Communicative psychoanalysis and sub-literal cognition. Journal of the American Academy of Psychoanalysis 27(3):471–502.

Haskell, R. E. (2000) Transfer of learning: Cognition, instruction, and reasoning. Academic Press: San Diego CA.

Haskell, R. E. (2002) Cognitive science and the origin of lexical LMS : A neurofunctional shift (NFS) hypothesis. Theoria et Historia Scientarium 6:291–326.

Haskell, R. E. (2003a) A logico-mathematic-structural methodology for the analysis and validation of sub-literal (SubLit) language and cognition. Journal of Mind and Behavior. Forthcoming

Haskell, R. E. (2003b) A meta level analysis of a logico-mathematic-structural methodology: Part I: Experimental design and epistemological issues. Journal of Mind and Behavior. Forthcoming

Haskell, R. E. (2003c) A meta level analysis of a logico-mathematic-structural methodology: Part II: A cognitive psychodynamics, masked priming, automatic schemata activation, and sub-literal (SubLit.) referents. Journal of Mind and Behavior. Forthcoming

Haskell, R. E. (2003d) The access paradox in analogical reasoning and transfer: Whither invariance? Theoria et Historia Scientarium. Forthcoming.

Haskell, R. E./Badalamenti, A. (in press) Algebraic structure of verbal narratives with dual meanings. Mathematics And Computer Modelling.

Hittmair-Delazer, M./Semenza, C./Denes, G. (1994) Concepts and facts in calculation. Brain 117:715–728.

Hittmair-Delazer, M./Sailer, U./Benke, T. (1995) Impaired arithmetic facts but intact conceptual knowledge: A single case study of dyscalculia. Cortex 31:139–147.

Jackendoff, R. (2002) Foundations of language: Brain, meaning, grammar, evolution. Oxford University Press: New York.

Kihlstrom, J. F. (1984) Conscious, subconscious, unconscious: A cognitive perspective. In: Bowers, K. S./Meichenbaum, D. (eds) The unconscious reconsidered. Wiley: New York, pp. 149–211.

Kihlstrom, J. F./Mulvaney, S./Tobias, B. A./Tobis, I. P. (2000) The emotional unconscious. In: Eich, E./Kihlstrom, J. F./Bower, G. H./Forgas, J. P./Niedenthal, P. M. (eds) Cognition and emotion. New York: Oxford University Press, pp. 30–86.

Koechlin, E./Dehaene, S./Mehler, J. (1997) Numerical transformations in five-month-old human infants. Mathematical Cognition 3(2):89–104.

Koehler, O. (1951) The ability of rats to count. Bulletin of Animal Behavior 9:41–45.

Kosslyn, S. M./Koenig, O. (1995) Wet mind: The new cognitive neuroscience. The Free Press: New York.

Lakoff, G./Johnson, M. (1980) Metaphors we live by. University of Chicago Press: Chicago.

MacCormac, E. R. (1985) A cogntive theory of metaphor. The MIT Press: Cambridge MA.

Marcus, G. F. (2001) The algebraic mind: Integrating connectionism and cognitive science. MIT Press: Cambridge.

Marr, D. (1983) Vision: A computational investigation into the human representation and processing of visual information. W. H. Freeman: New York.

Moyer, R. S./Landauer, T. K. (1967) Time required for judgements of numerical inequality. Nature 215(5109):1519–1520.

Naccache, L./Dehaene, S. (2001) Unconscious semantic priming extends to novel unseen stimuli. Cognition 80(3):215–229.

Norman, D. A. (1981) Categorization of action slips. Psychological Review 88:1–15.

Piaget, J. (1973) The affective unconscious and the cognitive unconscious. Journal of the American Psychoanalytic Association 21:249.

Reber, A. S. (1993) Implicit learning and tacit knowledge: An essay on the cognitive unconscious. Oxford University Press: New York.

Seyfarth, R. M./Cheney, D. L. (1984) Grooming, alliances, and reciprocal altruism in vervet monkies. Nature 308:541–43.

Smith, D. L. (1991) Hidden conversations: An introduction to communicative psychoanalysis. Tavistock/Routledge: London.

Smith, D. L. (1999) Freud's Philosophy of the unconscious. Kluwer Academic Publishers: Dordrecht.

Stanescu-Cosson, R./Pinel, P./Moortele, P. F./Le Bihan, D./Cohen, L./Dehaene, S. (2000) Understanding dissociations in dyscalculia. A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. Brain 123(11):2240–2255

Thompson, R. F./Mayers, K. F./Robertson, R. T./Patterson, C. J. (1970) Number coding in association cortex of the cat. Science 168:271–273.

Triesman, A. (1985) Preattentive processing in vision. Computer Vision, Graphics, and Image Processing 31:156–177.

Trivers, R. (1971) The evolution of reciprocal altruism. Quarterly Review of Biology 46:35–56.

Vaihinger, H. (1924) The philosophy of "as-if". Routledge and Kegan Paul: London. Originally published in 1911.

Vosniadou, S./Ortony, A. (eds) (1989) Similarity and analogical reasoning. Cambridge University Press: New York.

Whyte, L. L. (1978) The unconscious before Freud. Mentor Books: New York.

Wilkinson, G. W. (1984) Reciprocal food sharing in the vampire bat. Nature 308:181–84.

Wittenbrink, B./Judd, C. M./Park, B. (1997) Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. Journal of Personality and Social Psychology 72:262–274.

Woodruff, G./Premack, D. (1981) Primative [sic] mathematical concepts in the chimpanzee: Proportionality and numerosity Nature 293:568–570.

Wynn, K. (1992) Addition and subtraction by human infants. Nature 358:749–750.

Bence Nanay

Evolutionary Psychology and the Selectionist Model of Neural Development: A Combined Approach

Abstract

Evolutionary psychology and the selectionist theories of neural development are usually regarded as two unrelated theories addressing two logically distinct questions. The focus of evolutionary psychology is the phylogeny of the human mind, whereas the selectionist theories of neural development analyse the ontogeny of the mind. This paper will endeavour to combine these two approaches in the explanation of the human mind. Doing so might help in overcoming some of the criticisms of both theories. The first part of the paper mentions three standard objections to evolutionary psychology and then outlines three philosophical problems evolutionary psychology has to offer a solution to. The second part will try to show that an approach combining evolutionary psychology and the selectionist theory of neural development might overcome some of these objections.

Key words

Evolutionary psychology, neural development, selection, evolutionary explanations, developmental system theory.

What is Evolutionary Psychology?

The central claim of evolutionary psychology is that our mental capacities have to be analysed with reference to the environment in which they have evolved (BARKOW/COSMIDES/TOOBY 1992; BUSS 1994, 1995; DENNETT 1995; PINKER 1997; PLOTKIN 1997; WRIGHT 1994). Understanding why the human hand functions the way it does undoubtedly implies analysing the environment it has evolved in. The same could be said about mental capacities as well: the examination of the environment of our ancestors might help understand our present emotions or food preferences.

The most important point that has been made by evolutionary psychologists is that the environment our mental capacities have been adapted to is not necessarily the same as the environment we live in now.[1] To quote one of the best known examples: preference for sugar was adaptive in the Pleistocene environment where calorie-rich food was rare. In the present environment, however, the same preference is no longer adaptive, since it is not vital for survival any more (at least in some parts of the world) and it may also lead to obesity and bad teeth.[2] Our preference was fixed in the Pleistocene environment and it has not changed since then, but the environment itself has changed. Thus, in analysing a certain mental capacity, the evolutionary environment that has to be taken into consideration is not the present environment but rather the Pleistocene environment to which this mental capacity has been adapted. This environment is usually called the Environment of Evolutionary Adaptedness (EEA), and we do not have any direct evidence of what it looked like, but some of its characteristics can be postulated based on what we know about how our ancestors lived in the Pleistocene era.

Three Standard Objections to Evolutionary Psychology

Evolutionary psychology has been widely criticised for several reasons (see FODOR 1998 and GOULD 1997, for example). The three best known objections are those of innatism, adaptationism, and modularism. I summarise these standard objections briefly before turning to the ones I would like to pose.

Perhaps the most basic objection to the project of evolutionary psychology is related to one of the oldest topics in the philosophy of mind: the question of innatism.[3] Since mental capacities are analysed as being adapted to the Pleistocene environment, it has to be assumed that they are innate: they are genetically coded. Therefore, evolutionary psychology has to put the emphasis on the inborn characteristics of mental capacities rather than on the learned ones. Thus, so the objection goes, it cannot explain the acquired characteristics of our mental capacities. Steven PINKER argued powerfully against this objection claiming that innate and learned properties are not exclusive of each other; furthermore, the more innate apparatus we have, the more efficient our learning can be (PINKER 1997). I will discuss this point in more detail in Section III and Section V.

The second objection, that of adaptationism, also refers to an important debate in cognitive science, the adaptationism–exaptationism debate. Adaptationists claim that every evolutionary process is driven by natural selection. Therefore, in evolutionary explanations we have to suppose that every trait that has been formed by evolution is a result of natural selection; in other words, it is adaptive (DENNETT 1995; DAWKINS 1978; BARKOW/COSMIDES/TOOBY 1992). According to the alternative view, natural selection is the most important but not the only evolutionary mechanism. Some traits, for example, are by-products—or "exaptations"—of adaptive processes.[4] These by-products could gain adaptive function later but initially they were not adaptive at all. Evolutionary psychology is often accused of being adaptationist, since they describe mental capacities as adapted to a certain environment (to the Environment of Evolutionary Adaptedness).[5] In other words, for every mental capacity that is to be explained by evolutionary psychology there has to be an adaptationist story explaining why this mental capacity was adaptive for our ancestors in the Pleistocene environment. Even though the first theorists of evolutionary psychology followed this explanatory scheme, it is not at all necessary for the general approach of evolutionary psychology. PINKER, for example, analyses a number of mental capacities that are not adaptive themselves but they are indeed necessary consequences of other adaptive processes (PINKER 1997).

The third standard objection invokes the concept of modularism. Evolutionary psychology is said to imply modularism, that is, the claim that the mind consists of a number of separate modules that are isolated from each other; there is no interaction among them.[6] This seems to be a consequence of the methodology of evolutionary psychology, namely, of giving an evolutionary explanation for a certain mental capacity independently from all the other ones. The evolutionary explanation of mating preference, for example, is independent from the evolutionary explanation of colour vision (since different environmental challenges made them possible); therefore, the part of the brain responsible for mating preferences must be independent and isolated from the part of the brain responsible for colour vision. Although some evolutionary psychologists actually make this inference (MURPHY/STICH 2000; BARKOW/COSMIDES/TOOBY 1992) it has to be noted that this further step is not necessary and not very plausible either. It is very well possible that two mental capacities evolved independently, as a result of independent environmental effects, but there may still be interactions between the parts of the brain that are responsible for these two capacities (KARMILOFF-SMITH 1992). Thus, modularity is not a necessary feature of evolutionary psychology either.

Philosophical Problems with Evolutionary Psychology: Three New Challenges

The outlines of these three standard objections are obviously very sketchy and oversimplifying. Instead of examining these critical points, however, I would like to raise three new objections.

The first one concerns the question of the connection between the levels of *explanandum* and *explanans* of evolutionary psychology. The explained phenomena are mental entities: mating preferences, behaviour patterns, whereas the explanation refers to the genome; thus, a biologically plausible explanation has to provide a connection between these two levels of description.

The second objection is a more general concern about the separation of the genetically coded and the developmental elements in evolutionary explanations. This question is related to one of the standard objections mentioned above, namely, that evolutionary psychology is based on innatist assumptions. We have seen that evolutionary psychology does not necessarily imply that some mental capacities are entirely innate whereas some are entirely learned. As PINKER points out, there is a possibility of a non-innatist version of evolutionary psychology whereby learned characteristics are not necessarily excluded from its explanatory scheme. There are two problems with this proposal. The first is that PINKER does not say much about how this synthesis is supposed to work. The second problem, however, is more important. Even if we accept PINKER's depiction of mental capacities as partly genetically coded and partly learned, the explanatory scheme of evolutionary psychology can be used only to describe the innate component of a certain mental capacity. Therefore, another theory is needed to explain the role played by learning in this mental capacity. The problem is not only that evolutionary psychologists are not looking for such theories, but also the fact that an evolutionary explanation gained in this way would consist of two distinct parts: the evolutionary psychology story and the developmental one.

Such an explanatory scheme can be easily questioned from a biological point of view. One of the most influential approaches in recent evolutionary biology, the so called developmental system theory, emphasises that the genetically coded and the developmental part of evolutionary explanations cannot and should not be detached from each other (GRIFFITHS/GREY 1994; GODFREY-SMITH 2000; STERELNY/SMITH/DICKISON 1996). More precisely, they argue against a sharp distinction between "genetic" and "environmental" developmental causes. If their argument is correct, then evolutionary psychology has to integrate the explanation of the ontogeny of mental capacities. And no such account, be it either conceptual or empirical, has been offered so far.[7]

The third problem posed here concerns the evolution of the plasticity of the human mind. One of the most important features of the human mind is said to be its plasticity, that is, its ability to survive in various environments. The human mind is coded to survive not in one specific environment only, but in a number of very different environments. This is why humans with more or less identical genetic setups can be present everywhere from the Equator to the North Pole. Thus, the openness to various environmental effects seems to be an important characteristic of the human mind, and it would require evolutionary explanation. Evolutionary psychology, however, cannot give explanation for this, since our adaptation to a specific environment (the Environment of Evolutionary Adaptedness) cannot explain why we are able to survive in various other environments as well. Another kind of explanation is needed.

In the next sections, I give the outline of a theory that tries to combine evolutionary psychology and selectionist theories of neural development. My claim is that this combined theory can overcome the three objections mentioned here.

What are the Selectionist Theories of Neural Development?

The central claim of this paper is that the three objections outlined above can be overcome by combining evolutionary psychology with the selectionist theories of neural development. In order to do this, a brief summary of these selectionist theories has to be given.

According to the selectionist model of neural development, environmental effects select among our neural connections after birth: the connections that are used will survive, whereas the rest will die out (CHANGEUX 1985; ADAMS 1998). We are born with far more neural connections than we would need, and in the course of ontogeny some of these disappear, while others survive. This process itself resembles natural selection in several respects, since the phase of variation is followed by that of selection. I addressed the question of how similar or how different these two processes are elsewhere (NANAY 2001). For now, it is enough to mention that the selection of neural connections—if analysed in the light of recent evolutionary theories—cannot be explained sufficiently with the help of the model of natural selection (MAYNARD SMITH/SZATHMARY 1995; HULL 2001; HULL/LANGMAN/GLENN 2001).

I will not discuss one of the best known approaches among the selectionist theories of neural development, namely, the so-called neural DARWINISM. This approach was introduced by Gerald EDELMAN (EDELMAN 1987; EDELMAN 1990; EDELMAN/TONONI 2000) and its biological plausibility has often been questioned (the most influential criticism was given by CRICK 1989). In EDELMAN's theory, environmental effects select, not among single neural connections, but among neurone groups. For simplicity, I will focus on the environmental selection among neural connections and put EDELMAN's neurone groups aside.

The Proposal: Combining Evolutionary Psychology and Selectionist Theories of Neural Development

Selectionist theories of neural development explain the ontogeny of the human mind, whereas, as we have seen, evolutionary psychology explains its phylogeny. Perhaps that is why they tend to ignore each other: there are very few references to evolutionary psychology in the selectionist neural development literature while evolutionary psychology in turn ignores the selectionist view of neural development. What I want to suggest, however, is a combination of these two approaches. Evolutionary psychology is used to explain why we are born with a certain set of neural connections, and selectionist theories serve as explanations of how certain environmental effects select among these neural connections.

Note that this explanatory scheme uses two different environments to explain our present mental capacities. Firstly, it uses the postulated Pleistocene Environment of Evolutionary Adaptedness (needed by evolutionary psychology), and, secondly, the present environment that modifies the inborn set of neural connections (that is, selects among them). Rough adaptation to the Pleistocene environment is followed by the fine-tuning to the present one.

Answering the Three New Objections

My claim is that this combined approach can overcome the three objections to evolutionary psychology I have raised above. First of all, it can explain the plasticity of human mind, since different environmental effects will result in the survival of different sets of neural connections. Therefore, a more or less similar innate genetic setup can adjust to various environments (cf. DEACON 1997).

The challenge posed by developmental system theory, that is, the danger of the separation of genetically coded and developmental elements in evolutionary explanations, is more difficult to explain away and it requires some terminological groundmaking.

First of all, two central concepts of the philosophy of biology need to be introduced: replication and interaction. According to David HULL, selection consists of repeated cycles of replication and interaction (HULL/LANGMAN/GLENN 2001; HULL 1981). He analyses selection conceived traditionally as "heritable variation in fitness" as cycles of a copying process (replication) and the interaction with the environment (NANAY 2002; Cf. also GODFREY-SMITH 2000; BRANDON 1996).

In the most standard natural selection case the replicator is the gene, whereas the interactor is the organism itself. The genes are passed on and the organism interacts with the environment in such a way that this interaction causes the replication of genes to be differential. In other words, those genes that are responsible for the development of organisms that are more successful are more likely to replicate.

The developmental system theorists criticised the replication/interaction distinction, because they thought that according to the standard evolutionary explanation replication gives rise to innate properties, whereas interaction gives rise to acquired properties. According to this view, if we want to get rid of the innate/acquired dichotomy, the replication/interaction distinction also has to go. Their solution is to identify replicator with the entire life cycle. (GRIFFITHS/GRAY 1994)

The problem developmental system theory poses can be acknowledged without accepting their ultimate solution. I am reluctant to throw away the replication/interaction distinction (see NANAY 2002), but since every version of evolutionary psychology (even the so called evolutionary developmental psychology approach, see BJORKLUND/PELLEGRINI 2001) is vulnerable to the objection of developmental system theory, the challenge needs to be answered. I would like to argue that the combined approach can respond to this objection without discarding the replication/interaction distinction.

At this point it is necessary to examine how the replication/interaction distinction applies to the combined approach. The significant change compared to the standard approaches of evolutionary psychology is that the interaction between the organism and the environment is explained on the neural level: the environmental effects select among our neural connections after birth.

This, however, means that when analysing a property we do not need to decide whether its explanation needs reference to the replication or the interaction. Every single property must be analysed with reference to both. Everything that is in our mind came into being as a result of the environment's selection among the initial variety of neural connections. The initial variety of neural connections is genetically coded, that is, its explanation must include the analysis of the replication, whereas the environmental selection of these neural connection is the interaction between the organism and the environment.

As a result, the distinction between innate and learned characteristics is necessarily blurred. Every mental capacity is innate to some extent, since a set of neural connections (from which the relevant ones will be selected) is given genetically. On the other hand, every mental capacity is learned, to some extent, since the genetically coded set of neural connections undergoes a selective process after birth. Thus, there is a gradual transition between innate and learned mental capacities.

The third objection is probably even more difficult to answer. The *explanandum* and the *explanans* are still on different levels of description. The explained phenomena are behavioural patterns or food preferences, whereas the explanation ends with the selected neural connections. Note, however, that the original problem changes somewhat in the light of the combined approach. Now the two different levels of *explanandum* and *explanans* are not those of mental entities and genes, but mental entities and neural connections. This leads us, however, to a familiar problem of philosophy of mind, that is, to the problem of the relationship between mind and brain.

Answering the Three Standard Objections

Finally, it is important to examine how the combined approach outlined here can face the three standard objections to evolutionary psychology: those of innatism, adaptationism, and modularism. It has been shown that these objections apply to some approaches to evolutionary psychology but not to others. At this point it has to be examined whether the approach presented here is in the former category or not.

We have seen that in the combined approach the borderline between innate and learned mental properties is not clear-cut; there is a gradual transition between innate and learned characteristics. As we have seen, every mental capacity is innate to some extent, since a set of neural connections (from which the relevant ones will be selected) is given genetically, but at the same time, every mental capacity is learned, to some extent, since the genetically coded set of neural connections undergoes a selective process after birth. Thus, the combined approach is not only not innatist but denies the validity of the classical innate–learned (or nature–nurture) opposition as well.

Second, an approach combining evolutionary psychology and the selectionist theories of neural development does not imply adaptationism. This framework is equally suitable for both adaptationist and exaptationist explanations. The second objection was that some approaches of evolutionary psychology are adaptationist. The objection of adaptationism, however, seems orthogonal to the claims of the combined approach. The claim of standard evolutionary psychology is that the Environment of Evolutionary Adaptedness formed our mental capacities. What follows from the combined approach is that our mental capacities are formed by (1) the Environment of Evolutionary Adaptedness and (2) the present environment. Neither of these shaping processes are necessarily adaptations. It is very well possible to give an evolutionary explanation for the exaptation of a mental ability in this framework.

The third objection, namely, that of modularism, is more challenging for the combined approach. We have seen that in the explanatory scheme I am proposing, it is implicitly assumed that a certain neural structure should implement a certain mental capacity. A possible counterargument could be that this assumption sneaks in a basically modularist presupposition, since it requires the spatial identification of those neural connections that are responsible for the implementation of a certain mental capacity. And—the argument would continue—this leads to a methodology that isolates certain encapsulated modules of the brain and examines them independently from the rest of the brain.

This, however, is not necessarily so, since the assumption that it is worth analysing different mental capacities by examining different neural modules of the brain, does not necessarily imply the claim that the mind is built up from encapsulated modules that are isolated from each other. An interactionist theory must allow for the localisation of certain mental abilities. What interactionists oppose to is the supposition that the mind contains modules among which there are no connections. The combined view implies at least the possibility of the localisation of mental abilities in the brain, but it is neutral to the question of the existence of encapsulated modules.

Moreover, as KARMILOFF-SMITH points out, it is very likely that in the course of child development the mind is becoming more and more modular (KARMILOFF-SMITH 1992). If this is true, the "modularisation" (or compartmentalisation) of children's mind can be analysed well with the help of the selection among neural connections. Thus, the framework of the combined approach not only allows for either modularism or the ne-

Author's address

Bence Nanay, Department of Philosophy, University of California, Berkeley, 301 Moses Hall, Berkeley, CA. 94720, USA. Email: nanay@uclink.berkeley.edu

gation of it, but it also accommodates a theory that denies the modularism/interactionism distinction.

In conclusion, I would like to note that the argument presented here is just the skeleton of the approach that would try to combine evolutionary psychology and the selectionist theories of neural development. A great deal of empirical research has to be done in order to arrive at a detailed unified theory that includes both the phylogeny and the ontogeny of the human mind in one conceptual framework.

Acknowledgements

I am grateful to Frank SULLOWAY, Eors SZATHMARY and Felicitas BECKER for their comments on earlier versions of this paper.

Notes

1 This is one of the most significant differences between sociobiology and evolutionary psychology. Cf. WILSON 1975; LUMSDEN/WILSON 1981; WILSON 1978. For critical overviews see for example: KITCHER 1984; LEWONTIN/ROSE/KAMIN 1982.

2 This famous example was given by David BUSS. See BUSS (1995).

3 The question goes back to DESCARTES and LOCKE, but it is present in the recent cognitive science literature as well. See ELMAN et al. (1996), for example.

4 GOULD/VRBA (1982); GOULD (1996, 1997); GOULD/LEWONTIN (1979). In evolutionary biology, the introduction of the concept of exaptation was preceded by a somewhat similar debate on preadaptation. See GOULD (1996), DENNETT (1995). The question of adaptation in the case of the human mind is discussed by PINKER (1997) and PLOTKIN (1997).

5 The most important criticism on the adaptationist assumptions of evolutionary psychology was given by GOULD 1997.

6 On modularism in general see FODOR (1983), MITHEN (1996), KARMILOFF-SMITH (1992).

7 The so called evolutionary developmental psychology approach is no exception from this (see BJORKLUND/PELLEGRINI 2001; GEARY/BJORKLUND 2000; and BUSS 1999).

References

Adams, P. R. (1998) Hebb and Darwin. Journal of Theoretical Biology 195: 419–438.

Barkow, J. H./Cosmides, L. M./Tooby, J. (eds) (1992) The adapted mind: Evolutionary psychology and the generation of culture. Oxford University Press: New York.

Bjorklund, D. F./Pellegrini, A. D. (2001) Evolutionary Developmental Psychology. APA Press: Washington DC.

Brandon, R. N. (1996) Concepts and Methods in Evolutionary Biology. Cambridge University Press: Cambridge. Buss, D. M. (1994) The evolution of desire: Strategies of human mating. Basic Books: New York.

Buss, D. M. (1995) Evolutionary psychology: A new paradigm for psychological science. Psychological Inquiry 6:1–30.

Buss, D. M. (1999) Evolutionary psychology: The new science of the mind. Allyn and Bacon: Boston.

Changeux, J.-P. (1985) Neuronal man: The Biology of mind. Pantheon: New York.

Crick, F. (1989) Neural Edelmanism. Trends in Neuroscience 12:240–248.

Dawkins, R. (1986) The selfish gene. Oxford University Press: Oxford.

Deacon, T. (1997) The symbolic species. W. W. Norton: New York.

Dennett, D. C. (1995) Darwin's dangerous idea. Touchstone: New York.

Edelman, G. M. (1987) Neural Darwinism: The theory of neuronal group selection. Basic Books: New York.

Edelman, G. M. (1990) The remembered present: A biological theory of consciousness. Basic Books: New York.

Edelman, G. M./Tononi, G. (2000) A universe of consciousness. Perseus Books: New York.

Elman, J. L./Bates, E. A./Johnson, M. H./Karmiloff-Smith, A./Parisi, D./Plunkett, K. (1996) Rethinking innateness: A connectionist perspective on development. MIT Press: Cambridge MA.

Fodor, J. A. (1983) The modularity of mind. MIT Press: Cambridge MA.

Fodor, J. A. (1998) The trouble with psychological darwinism. London Review of Books 20(2):4–9.

Geary, D. C./Bjorklund, D. F. (2000) Evolutionary developmental psychology. Child Development 71:57–85.

Godfrey-Smith, P. (2000) The replicator in retrospect. Biology and Philosophy 15:403–423.

Gould, S. J. (1996) Full House. Crown: New York.

Gould, S. J. (1997) Evolution: The pleasures of pluralism. New York Review of Books June 26:47-52.

Gould, S. J./Lewontin, R. (1979) The sprandels of San Marco and the panglossian paradigm/ Proceedings of the Royal Society B205:581–598.

Gould, S. J./Vrba, E. S. (1982) Exaptation: A missing term in the science of form. Paleobiology 8:4–15.

Griffiths, P./Gray, R. (1994) Developmental systems and evolutionary explanation. Journal of Philosophy 91:277–304.

Hull, D. L. (1981) Units of evolution: A metaphysical essay. In: Jensen, U. J./Harré, R. (eds) The philosophy of evolution. Harvester Press: Brighton, pp. 23–44.

Hull, D. L. (2001) Science and selection. Cambridge University Press: Cambridge.

Hull, D. L./Langman, R. E./Glenn, S. S. (2001) A general account of selection: Biology, immunology and behavior. Behavioral and Brain Sciences 24:511–528.

Karmiloff-Smith, A. (1992) Beyond modularity. MIT Press: Cambridge MA.

Kitcher, P. (1982) Abusing science. MIT Press: Cambridge

MA.
Lewontin, R./Rose, S./Kamin, L. (1984) Not in our Genes: Biology, ideology and human nature. Pantheon: New York.
Lumsden, C. J./Wilson, E. O. (1981) Genes, mind, and culture: The coevolutionary process. Harvard University Press: Cambridge MA.
Maynard Smith, J./Szathmary, E. (1995) Major transitions of evolution. Oxford University Press: Oxford.
Mithen, S. (1996) The prehistory of the modern mind. Phoenix: London.
Murphy, D./Stich, S. (2000) Darwin in the madhouse: Evolutionary psychology and the classification of mental disorders. In: Carruthers, P./Chamberlain, A. (eds) Evolution and the Human Mind. Cambridge University Press: Cambridge: pp. 62–92.
Nanay, B. (2001) A more pluralist typology of selection processes. Behavioral and Brain Sciences 24:547–548.
Nanay, B. (2002) The return of the replicator: What is philosophically significant in a general account of replication and selection? Biology and Philosophy 17:109–121.
Pinker, S. (1997) How the mind works. Norton: New York.
Plotkin, H. C. (1997) Evolution in mind. Allen Lane: New York.
Sterelny, K./Smith, K./Dickison, M. (1996) The extended replicator. Biology and Philosophy 11:377–403.
Wilson, E. O. (1975) Sociobiology: A new synthesis. Harvard University Press: Cambridge MA.
Wilson, E. O. (1978) On human nature. Harvard University Press: Cambridge MA.
Wright, R. (1994) The moral animal: Why we are the way we are: The new science of evolutionary psychology. Random House: New York.

Steven M. Platek

Unconscious Reactions to Children's Faces: The Effect of Resemblance

Abstract

When it comes to positive reactions toward children's faces resemblance matters more for males than for females. A morphing experiment was conducted to test whether actual resemblance (i.e., the degree to which a child's face shares characteristics with a person) or social mirror-mediated resemblance (i.e., being told a child's face looks like you) accounted for more variance in reactions towards a child's face. Social mirror information affected both males and females in a similar fashion, but males were more likely than females in all conditions to react positively towards faces based on whether they shared actual characteristics with the face. In a re-analysis of existing data, no subject reported being aware/conscious of the effects of resemblance on their decisions about how to hypothetically invest in the faces.

Key words

Parental investment, paternal resemblance, cuckoldry, paternal uncertainty, social mirror, sex differences, facial resemblance, unconscious perceptions.

Introduction

Species that utilize internal fertilization (e.g., humans) are faced with an asymmetry of parental certainty; i.e., unlike females who can be 100% certain that they share genes in common with their offspring, males can never be certain of paternity. Parental investment in such species often parallels this paternal uncertainty; i.e., males invest less in offspring than females and male investment decisions are mediated in part by the male's certainty of paternity. There are three ways in which a male can increase his level of certainty that he is the source of paternity: (1) he can monitor/mate guard and/or sequester the female partner during the period that she is fertile (DALY/WILSON 1998), (2) he can rely on intra-vaginal inter-male competition such as sperm competition (BIRKHEAD 1995, 1996; SCHAKELFORD et al. 2002) or semen displacement (GALLUP et al. in press), or (3) he can attempt to assess paternity post-parturition. We (PLATEK et al. 2002, in press) and others (e.g., DALY/WILSON 1998) have hypothesized males infer paternity by assessing physical resemblance.

An indirect way that resemblance might affect male behavior toward children is by a social mirror (BURCH/GALLUP 2000); i.e., other people can ascribe paternal resemblance to the child as a means of swaying a male to act more positively towards the child. DALY/WILSON (1982) and REGALSKI/GAULIN (1993) have shown that mothers and family members actively ascribe paternal resemblance to children and that when males express doubt they are quick to reassure them of resemblance. Males may be predisposed to take into account social mirror information because of the importance of paternity (see NEFF/SHERMAN 2002 for model of parentage).

BURCH/GALLUP (2000) found that the more a sample of convicted spouse abusers felt that their children looked like them the better the children were treated. The childhoods of the abusive males themselves were also affected by how much they thought they resembled their fathers. Perceptions of paternal resemblance were correlated with the incidence of physical and sexual abuse they experienced as children, as well as feelings of closeness to their father. How often others had told them that the child physically resembled them also correlated with his ratings of his relationship to the child. Social affirmation of resemblance between the subject and his father also correlated with the incidence of reported childhood abuse.

A male, however, may also assess the degree to which a child actually resembles him. PLATEK et al. (2002) morphed the faces of participants with the face of a toddler and measured reactions to hypothetical investment questions (e.g., which one of these children would you spend the most time with? Which one would you adopt?). Males were more likely than females to select a face they had been morphed with when asked to react positively towards the faces. Thus it seems that actual resemblance also plays a role in a male's reactions toward children's faces and this might be modulated by a mechanism/module (e.g., self-referent phenotype matching; see LACY/SHERMAN 1983; NEFF/SHERMAN 2002) dedicated to controlling the affective nature of males' reactions toward children. The interaction of actual resemblance and social mirror-mediated resemblance has yet to be systematically tested.

It is obvious that convincing a male of paternity and securing his investment would be in the evolutionary best interests of females. However, because the incidence of cuckoldry is appreciable (1–20%, see BAKER/BELLIS 1995; CERDA-FLORES et al. 1999; SASSE et al. 1994; SYKES/IRVEN 2000) it is hardly in the best interests of males to be easily convinced. If ascriptions of resemblance were completely persuasive throughout evolutionary history, males might have been deceived in investing in children that were the byproduct of cuckoldry. Those who remained wary and used their own perceptions of resemblance and invested accordingly might have stood a better chance of passing on their genetic material and maximizing their fitness through their reproductive and parental efforts.

MCLAIN et al. (2000) investigated this idea by objectively comparing new mothers' ascriptions of paternal resemblance to the ability of independent raters to match photos of children to their putative fathers. Maternal ascriptions of resemblance could not be verified by the objective, unrelated raters. In no case were the infants' pictures matched to the fathers' photos more often than chance; the mothers' opinions, although highly reliable, held no validity. These findings would, in turn, explain the reluctance in the males to prematurely agree with their partner's assertions of paternity (DALY/WILSON 1982; REGALSKI/GAULIN 1993). In some cases these males would only agree after several maternal attempts to persuade them. Interesting as these data are, they are still flawed in that actual paternity was never determined, which could have masked/obscured independent raters' ability to match children to the males (see also BRÉDART/FRENCH 1999; CHRISTENFELD/HILL 1995; NESSE/SILVERMAN/BORTZ 1991).

In addition, males may have adopted a strategy of comparing their offspring to their kin in order to assess resemblance and also choose to believe information provided only by those that also shared genes in common with him (i.e., resembled him). PLATEK et al. (in press) and DEBRUINE (2002) have provided indirect evidence that parentage and trust may also be mediated by kin resemblance. PLATEK et al. (in press) found that males would react favorably toward children's faces that shared 25% of their characteristics, which is approximately the proportion of genes shared in common with kin such as grandchildren, nieces and nephews, aunts and uncles, and half-siblings. DEBRUINE (2002) found that participants trusted those faces that they shared characteristics with.

Although it appears adaptive for both females to ascribe resemblance and for males to be suspicious, it remains to be seen just how this "social mirror" interacts with actual resemblance in having effects on male investment strategies and reactions towards children. This study was designed to examine how actual resemblance and social mirror-mediated resemblance interact using a computer morphing paradigm (PLATEK et al. 2002, in press).

Methods

Subjects

Forty (20 male, 20 female) undergraduates were recruited from the State University of New York at Albany as subjects and received extra credit for their participation. Subjects were informed ahead of time that they would be participating in a study that required having their picture taken and that they were going to be asked questions about childcare. This study was approved by the local institutional review board and all subjects gave written informed consent.

Pictures

Pictures of subjects were taken using a Hewlett Packard Model 315 digital camera. Subjects were asked not to smile or frown for the picture and to try to maintain a neutral unexpressive face. The images were processed using a 233mHz PC and Paint Shop Pro (Version 3.11) software. All images were converted to 256 shades of gray, cropped just under the chin, from ear to ear, and just below the hairline, and images were centered. All images were made the same size (200 X 150 pixels in Paint Shop Pro, on screen pictures were represented by a 4.125:3 width to height ratio) and brightness using the color editor and histogram rectification properties in Paint Shop Pro. The images were mounted on a white background.

Each subject's picture was morphed (Ulead Morph Editor version 1.0) with the image of a female and a male 2-year-old child so that the stimulus image was 50% of the subject's face and 50% of one of the children's faces (see PLATEK et al. 2002 for detailed description of methodology).

Design

The entire experiment was computerized using SuperLab (Cedrus, version 2.01). Each trial consisted of the subject being shown 5 faces in a semi-circular array with a question (e.g., "Which of these children would you spend the most time with?", "Which one of these children would you be most likely to adopt?") embedded in the center of the array (see PLATEK et al. 2002). There were six types of arrays. To replicate our previous findings (PLATEK et al. 2002, in press), two arrays consisted of only morphed faces; i.e., no social mirror information was provided (one array condition contained a face that had the subject's face morphed and one did not as a control, both were randomized as to whether they showed faces that were morphed with the face of the little girl or boy). The other four arrays (social mirror conditions) consisted of subject-child face morphs, a question, and social mirror feedback generated by the computer.

The subjects were told that the computer compared their picture to a database of pictures and generated feedback as to whether any of the children's faces in a particular array shared facial characteristics with them. In social mirror Condition 1 self-morphs were labeled as resembling the subject ("This One Does" appeared above the self-morph face). In social mirror Condition 2 self-morphs were labeled as not resembling the subject ("Does Not" appeared above the self-morph face). In social mirror Conditions 3 and 4 other-morphs were labeled as resembling (Condition 3) or not resembling (Condition 4). Social mirror labels for Conditions 3 and 4 were paired with a randomly selected other morph.

The six face array conditions were paired with each of the 10 questions (see Table 1) once for a total of 60 stimulus trials; 6 conditions, 10 questions per trial. Whether the subject saw the faces morphed with the girl or boy child's face in any condition was randomized across subjects, conditions, and questions. The coordinates in which a face was presented was also randomized across trials; i.e., the placement of a face in an array was changed across trials. Latency to respond was recorded for all questions by SuperLab (Cedrus: Version 10.4) as the latency between the onset of an array and a subject's response. Subjects responded by selecting letters on the computer keyboard that corresponded to letters presented under the faces on the screen and after they selected one face for an array the screen was erased and a new array and question was presented. At the end of the experiment subjects were asked how they made their choices (open-ended) and how difficult it was to make a choice on a scale of 1–5 (1 being the easiest and 5 being the most difficult).

	Males (N = 20)	Female (N = 20)
Which one of these children would you adopt?	80%*	30%
Which one of these children do you find to be the most attractive (cute)?	60%	40%
Which one of these children would you be most comfortable spending time with?	90%*	40%
Which one of these children would you spend the least time with?	10%	10%
Which one of these children would you spend US$50 on if you could only spend it on one child?	70%	50%
Which one of these children would you spend US$50 on last?	70%	70%
If one of these children damaged something valuable of yours, which one would you punish MOST?	0%	0%
If one of these children damaged something valuable of yours, which one would you punish the least?	0%	0%
If you were forced to pay child support to these children, which would you resent MOST having to pay for?	10%	0%
If you were forced to pay child support to these children, which would you resent LEAST having to pay for?	30%	20%

Table 1. Sex differences in selecting self-child-morphs in response to hypothetical investment questions (* $p < 0.05$).

Results

There were no main effects for the sex of the toddler face; i.e., both within and between subjects whether a person's face was morphed with the face of a male or female child did not make a difference in terms of reactions towards the faces.

Selection of faces: Non-social mirror conditions

When given no social mirror information, males were more likely than chance (binomial test) to select a face they had been morphed with in response to the questions: "Which one of these children would you adopt?", $p < 0.001$, and "Which one of these children would you be most comfortable spending time with?", $p < 0.05$. Females were no more likely than chance to choose the face they had been morphed with in response to any question. When given no "social mirror" feedback males were more likely to choose faces they had been morphed with than were females for positive-type questions (see Table 1). FISHER's exact probability test showed that males were more likely to choose their own face morph in response to "Which one of these children would you be most likely to adopt?", $p < 0.05$ and "Which one of these children do you think that you would spend the most time with?", $p = 0.05$. Unlike males, there were no questions where females were more likely to choose their own face-toddler morph. In Condition 2, where the subject's face was not morphed with any of the child faces, there were no sex differences in the likelihood to select any particular face, and neither males nor females chose any face more often than chance.

"Social mirror" conditions: Chance selection of faces

Social mirror Condition 1: Self-morph labeled as resembling. Binomial tests revealed that males chose the face they had been morphed with more often than chance in response to the questions: "Which one of these children would you adopt?", $p < 0.05$, "Which one of these children would be most comfortable spending time with?", $p < 0.05$, "Which one of these children do you find to be the most attractive (cutest)?", $p < 0.05$, and "If you were forced to pay child support to these children which child would you resent least having to pay for?", $p < 0.01$. Females did not select the face they had been morphed with more often than chance in response to any of the questions. FISHER's exact probability test showed that males chose their self-morph more often than females in response to the questions: "Which one of these children would you adopt?", $p < 0.01$, and "Which one of these children would you be comfortable spending the most time with?", $p < 0.05$. (See Table 2).

Social mirror Condition 2: Self-morph labeled as not resembling. Males selected a face they had been morphed with more often than chance (binomial test) in response to the questions: "Which one of these children would you adopt?", $p < 0.01$, "Which one of these children would you be most comfortable spending time with?", $p < 0.01$, "If you were forced to pay child support to these children, which one would you resent least having to pay for?", $p < 0.05$, and "Which one of these children would you give \$50 to first?", $p < 0.05$. Females did not choose their self-morph more often than chance in response to any question. FISHER's exact probability test revealed that males selected a face they had been morphed with more often than chance in response to the questions: "Which one of these children would you adopt?", $p < 0.05$, "Which one of these children would you be most comfortable spending time with?", $p < 0.05$, "If you were forced to pay child support to these children, which one would you least resent having to pay for?", $p < 0.05$, and "Which one of these children would you give \$50 to first?", $p < 0.01$. (See Table 2a).

Social mirror Condition 3: Other-morph labeled as resembling. Males were more likely than chance (binomial test) to select the face they had been morphed with in response to the questions: "Which one of these children would you adopt?", $p < 0.01$, "Which one of these children would you be most comfortable spending time with?", $p < 0.05$, "Which one of these children do you find to be the most attractive (cutest)?", $p < 0.05$, "If you were forced to pay child support to these children, which one would you least resent having to pay for?", 0.05, and "Which child would you give \$50 to first?", $p < 0.05$. Females did not select a face they had been morphed with more likely than chance for any question in this condition. FISHER's exact probability test revealed that males selected a face they had been morphed with more often than females in response to the questions: "Which one of these children would you adopt?", $p < 0.05$, "Which one of these children do you find to be the most attractive (cutest)?", $p < 0.01$, and "Which one of these children

Question	Condition 1 Male	Condition 1 Female	Condition 2 Male	Condition 2 Female	Condition 3 Male	Condition 3 Female	Condition 4 Male	Condition 4 Female
Adopt	*80% (.012)	10%	*85% (.003)	45%	*85% (.003)	45%	*80% (.012)	55%
Spend Time	70%	40%	*85% (.003)	45%	*80% (.012)	60%	*80% (.012)	65%
Comfortable	*80% (.012)	40%	70%	40%	*75% (.041)	45%	*85% (.003)	55%
Attractive	*80% (0.12)	35%	20%	10%	*85% (.003)	35%	*80% (.012)	75% (.041)
$50 last	5%	0%	10%	0%	0%	15%	0%	0%
Punish Most	0%	5%	0%	25%	5%	15%	0%	0%
Punish Least	50%	65%	70%	30%	65%	50%	65%	0%
Resent most For child Support	0%	0%	65%	55%	55%	65%	50%	70%
Resent least For child Support	*85% (.003)	65%	*75% (.041)	30%	*75% (.041)	50%	*75% (.041)	35%
$50 First	70%	45%	*75% (.041)	10%	*80% (.012)	40%	*100% (.001)	35%

Table 2. Percentage of time and significance level of the binomial probability that a self-morph was selected for each question for each condition more often than chance (.20) under conditions in which the child morphs were accompanied with social mirror information (* $p < 0.05$).

would you give $50 to first?", $p < 0.05$ (see Table 2). Specific analyses of whether males were drawn more to selecting the other morph face labeled as resembling them (i.e., social mirror) more often than chance revealed that males were relatively unaffected by this manipulation; i.e., it was no more likely than chance for males to choose this face than it was for them to choose any of the other other-morphs.

Social mirror Condition 4: Other-morph labeled as not resembling. Males were more likely than chance to select a face they had been morphed with in response to the questions "Which one of these children would you give $50 to first?", $p < 0.05$, "Which one of these children would you be most comfortable spending time with?", $p < 0.05$, "Which one of these children do you find to be the most attractive (cutest)?", $p < 0.01$, "If one of these children damaged something of yours, which one would you punish least severely?", $p < 0.05$, and "Which one of these children would you give $50 to first?", $p < 0.01$. In this condition only, females were more likely than chance to select a face they had been morphed with in response to the question "Which one of these children do you find to be the most attractive (cutest)?", $p < 0.05$. (See Table 2).

When collapsing across all social mirror conditions, males were more likely than females to select a face that they had been morphed with in response to the questions: "Which one of these children would you adopt?", $p < 0.01$, "Which one of these children do you find to be the most attractive (cutest)?", $p < 0.05$, "If you were forced to pay child support to these children, which one would you resent least having to pay for?", $p < 0.01$, and "Which one of these children would you give $50 to first?", $p < 0.01$.

Social mirror condition				
Gender	Condition 1	Condition 2	Condition 3	Condition 4
Male	2.90 (± 1.71)	1.35 (± 0.98)	2.60 (± 2.56)	3.45 (± 0.944)
Female	1.59 (± 1.76)	0.32 (± 1.46)	1.14 (± 1.83)	2.00 (± 1.74)

Table 3. Mean (± S.D.) favorable reactions towards self-morphs across social mirror conditions. For both males and females social mirror condition had a significant effect on favorable reactions towards self-morphs.

Social mirror condition: Mean rank				
Gender	Condition 1	Condition 2	Condition 3	Condition 4
Male	26.10	26.02	27.40	27.33
Sex Difference	($p < 0.014$)	($p < 0.018$)	($p < 0.002$)	($p < 0.002$)
Female	17.32	17.39	16.14	16.20

Table 4. A Mann-Whitney U test revealed sex difference in favorable reactions towards self-morphs across social mirror conditions.

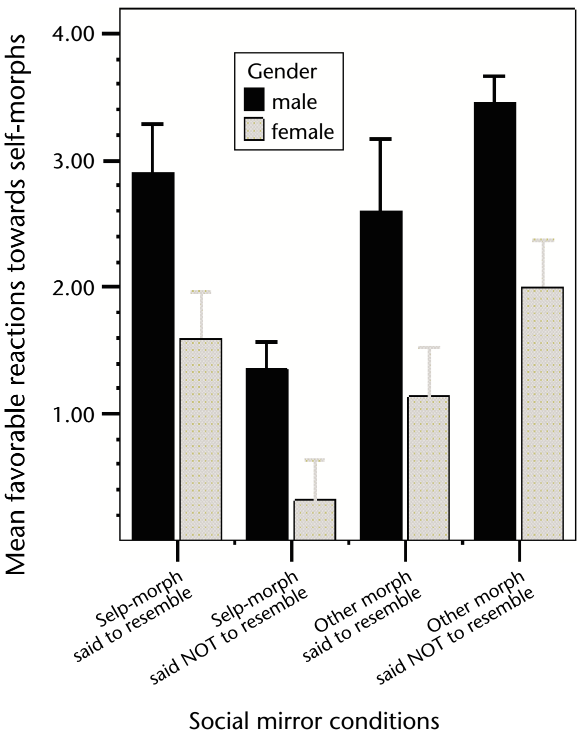

Figure 1. Sex differences in favorable reactions toward self-morphs across the four social mirror conditions (composite scores range: 1–5).

"Social mirror" conditions: Actual resemblance vs. social mirror-mediated resemblance

In order to compare the effect that social mirror information had on reactions toward children's faces, we created favorable response to self-morph scores by subtracting the number of times a subject selected a self-morph in response to a negative question (e.g., which one would you spend the least amount of time with?) from the number of times that a subject selected a self-morph in response to a positive question (e.g., which one of these children would you adopt?). These scores were used to calculate overall differences between social mirror conditions and sex differences for each condition and across all four conditions.

A FRIEDMAN's (non-parametric) analysis showed that for both males and females social mirror information had an effect on favorable responses to self-morphs, $p < 0.01$, (see Table 3 and Figure 1). Subjects selected self-morphs the least when their self-morph was labeled as not resembling them. A MANN-WHITNEY U test for each social mirror condition however, revealed that males relied more heavily on actual resemblance in all social mirror conditions (see Table 4 and Figure 1).

Difficulty and latency to respond

Women ($M = 3.95$) found it more difficult to respond to the faces than men ($M = 2.55$), $F(1, 38) = 9.069$, $p < 0.05$. There were no differences between genders in latency to respond to any questions and there was no statistical consistency in reports of how people made their choices. However, females often tried to try to generate a personality for the faces; i.e., they would report selecting a face that looked like it "wanted to be adopted", or that appeared to "need attention". Males did not respond in this way. No subject reported resemblance as a factor in how they chose a face.

In a test of how actual resemblance and social mirror-mediated resemblance interact to effect reactions toward computer generated children's faces, we found that social mirror information affected both male and female reactions similarly, but unlike females, males selected a face primarily based on actual resemblance. This study replicates our previous findings that males utilize actual resemblance in their reactions toward hypothetical children (PLATEK et al. 2002, in press) and supports the idea that social-perceptions of resemblance affect father–child relationships (BURCH/GALLUP 2000).

Discussion

In this experiment, the only context in which social affirmation seemed to affect reactions towards the faces were when the subject was told that a self-morph did not resemble him and that a non-self-morph did resemble him (see Figure 1). These conditions represent a situation in which sensory information about resemblance is being weighed against the social mirror feedback the male is receiving. Our data suggest that when social mirror information is available it can affect reactions towards a child's face when the information between the actual and the social mirror-mediated resemblance is incongruous; i.e., when a self-morph is said not to resemble and a non-self-morph is said to resemble, faces that actually resemble subjects are selected less. These incongruous social mirror conditions might represent instances of cognitive interference in information processing, thus affecting optimal functioning of a resemblance detection module.

As in previous studies (PLATEK et al. 2002, in press) none of the subjects were aware of the effect resemblance had on their choices. In a re-analysis of existing data from our previous morphing studies (PLATEK et al. 2002, in press) in conjunction with the present data, no subject out of 170 reported using resemblance to choose a child's face. When queried about their choices, none of the subjects identified resemblance as a factor in how they chose which child to select, nor did they realize that their faces had been morphed with the child. In fact during debriefing when subjects were told the hypothesis, and shown the morphing procedure, most subjects responded with surprise and asked to see the faces again in an attempt to identify more consciously which face it was that they had been morphed with. Even under these conditions subjects still had difficulty selecting the face they had been morphed with. It was not until their real picture was available for comparison to the child morphs that subjects could tell which face they had been morphed with.

The relationship between social mirror-mediated resemblance and actual resemblance is complex. Actual and social mirror mediated resemblance might fit a model developed by NEFF/SHERMAN (2002; see also HAUBER/SHERMAN 2001). According to this model actual resemblance may act as a "cue" and social mirror-mediated resemblance may act as a "predisposition". If physical resemblance is a cue, as defined by NEFF and SHERMAN, it might be possible for it to override the predisposition to trust social mirror information, but the two can still interact additively in persuading a male to invest in offspring appropriately. Additionally, there should be individual variation in the way in which the two types of information interact.

In conclusion, social mirror ascriptions of resemblance affect both male and female reactions towards the faces of children; however males rely more heavily on actual resemblance than females in their reactions to children's faces.

Author's address

Steven M. Platek, Department of Psychology, Drexel University, 1505 Race Street, Mail Stop 626, 10th Floor Bellet Bldg., Philadelphia, PA 19104, USA.
Email: steven.m.platek@drexel.edu

Acknowledgements

The author thanks Ivan PANYAVIN, Thomas MYERS, Samuel CRITTON, Julian Paul KEENAN, Sid O'BRYANT, Carlos FINLAY, Susan HUGHES, Marissa HARRISON, Brett WASSERMAN, David FREDERICK, Chris ANDERSON, Anthony, Michelle, and Joseph RUBINO, and Patricia and Sabrina SCOTT for their assistance with design materials. The author also thanks Gordon GALLUP, Jr. and Rebecca BURCH for helpful comments on an earlier version of this manuscript.

References

Baker, R. R./Bellis, M. A. (1995) Human sperm competition: Copulation, masturbation, and infidelity. Chapman and Hall: London.

Birkhead, T. R. (1995) Sperm competition: Evolutionary causes and consequences. Reproduction, Fertility, and Development 7:755–775.

Birkhead, T. R. (1996) Sperm competition: Evolution and mechanisms. Current Topics in Developmental Biology 33:103–158.

Brédart, S./French, R. (1999) Do babies resemble their fathers more than their mothers? A failure to replicate Christenfeld and Hill. Evolution and Human Behavior 20:129–135.

Burch, R. L./Gallup, G. G. Jr. (2000) Perceptions of paternal resemblance predict family violence. Evolution and Human Behavior 21:429–435.

Cerda-Flores, R. M./Barton, S. A./Marty-Gonzales, L. F/Rivas, F./Chakraborty, R. (1999) Estimation of nonpaternity in the Mexican population of Nuevo Leon: A validation study with blood group markers. American Journal of Physical Anthropology 109:281–293.

Christenfeld, N./Hill, E. (1995) Whose baby are you? Nature 378:669.

Daly, M./Wilson, M. (1982) Whom are newborn babies said to resemble? Ethology and Sociobiology 3:69–78.

Daly, M./Wilson, M. (1998) The truth about Cinderella: A Darwinian view of parental love. Yale University Press: New Haven.

DeBruine, L. M. (2002) Facial resemblance enhances trust. Proceedings of the Royal Society London B. Biological Science 269:1307–1312.

Gallup, G. G. Jr./Burch, R. L./Zappieri, M. L./Parvez, R./Stockwell, M. (in press) The penis as a semen displacement device. Evolution and Human Behavior. In Press.

Hauber, M. E./Sherman, P. W. (2001) Self-referent phenotype matching: Theoretical considerations and empirical evidence. Trends in Cognitive Sciences 10:609–616.

Lacy, R. C./Sherman, P. W. (1983) Kin recognition by phenotype matching. American Naturalist 121:489–512.

McLain, D. K./Setters, D./Moulton, M. P./Pratt, A. E. (2000) Ascription of resemblance of newborns by parents and nonrelatives. Evolution and Human Behavior 21:11–23.

Neff, B. D./Sherman, P. W. (2002) Decision making and recognition mechanisms. Proceedings of the Royal Society of London (B) 269:1435–1441.

Nesse, R./Silverman, A./Bortz, A. (1990) Sex differences in ability to recognize family resemblance. Ethology and Sociobiology 11:11–21.

Platek, S. M./Burch, R. L./Panyavin, I. S./Wasserman, B. H./Gallup, G. G. Jr. (2002) Reactions toward children's faces: Resemblance matters more for males than females. Evolution and Human Behavior 23:159–166.

Platek, S. M./Critton, S. R./Burch, R. L./Frederick, D. A./Myers, T. E./Gallup, G. G. Jr. (in press) How much paternal resemblance is enough? Sex differences in effects of resemblance but not in the detection of resemblance. Evolution and Human Behavior.

Regalski, J./Gaulin, S. (1993) Whom are Mexican infants said to resemble? Monitoring and fostering paternal confidence in the Yucatan. Ethology and Sociobiology 14:97–113.

Sasse, G./Muller, H./Chakrabroty, R./Ott, J. (1994) Estimating the frequency of non-paternity in Switzerland. Human Heredity 44:337–343.

Schakelford, T. K./LeBlanc, G. J./Weekes-Schakelford, V. A./Bleske-Rechek, A. L./Euler, H. A./Hoier, S. (2002) Psychological adaptations to human sperm competition. Evolution and Human Behavior 23:123–138.

Sykes, B./Irven, C. (2000) Surnames and the Y chromosome. American Journal of Human Genetics 66:1417–1419.

David Livingstone Smith

A Breast of Flesh Air

The Evolution of Unconscious Verbal Communication

Abstract

Deception is widespread amongst organisms, including human beings. According to neo-DARWINIAN theory the human capacity for self-deception emerged from a coevolutionary arms race between the capacity for perpetrating intraspecific tactical deception and the cognitive aptitude for penetrating it. By hiding our true intentions from ourselves, we are better able to conceal them from others. I suggest that the evolution of self-deception has created selection pressures that have given rise to sophisticated forms of unconscious social cognition and communication. Human beings unconsciously scan one another's behavior for subtle signs of deception. This information is unconsciously analyzed and the perceptions are conveyed by means of verbal communications that carry unconscious or "sub-literal" meanings.

Key words

Evolution, unconscious, conscious, self-deception, gossip.

Introduction

One hot summer evening a man at a party decides to step outside to cool off. In the hallway he comes face to face with an attractive woman wearing a low-cut dress. Embarrassed, he mumbles to her "Pardon me, I need to get a breast of flesh air".

Verbal parapraxes or "FREUDIAN slips" are almost always attributed to some sort of mechanical failure within the mind. From FREUD onwards, cognitive scientists have understood verbal slips as *errors* in speech production brought about by competing mental representations. From an evolutionary perspective, features that might superficially seem bizarre or pathological may have adaptive value. Evolutionary psychologists should entertain the possibility that speech errors *may* be adaptive communicative acts, designed to influence the social environment in specific ways advantageous to the slipper.

Let us consider the opening vignette. In committing his "error" the protagonist obliquely communicates to the woman that he is sexually aroused by her ("I need to get a breast of flesh"). Instead of overtly saying "I am aroused by your breasts", he makes the point cryptically. What advantage might there be in communicating sexual desire in this manner? The gain is perhaps the same as that gained by intentionally cryptic flirtations in the choreography of human mating. We say things, and do things, that can be interpreted as suggestive of sexual interest in order to assess the responsiveness or non-responsiveness of potential mates while minimizing costs (e.g., cost to one's reputation). In this instance there is insufficient information to push the analysis any further. My purpose is not so much to provide a convincing explanation as it is to suggest an alternative to conventional ways of thinking about verbal communication.

Unlike the disruptive slip, the *double entendre* seamlessly fuses two disparate semantic representations in a single utterance.

Mae West: How big are you fella?
Cary Grant: Six foot six inches.
Mae West: Well let's forget about the six feet and concentrate on the six inches.

Double entendres make use of what might be called, using an analogy from physics, the principle of semantic superposition. Imagine a sentence as a vibrating string, with the amplitude and frequency of the vibrations standing for the semantic content of the sentence. Just as the string simultaneously vibrates on both horizontal and vertical dimensions, so the sentence may simultaneously encode two separate semantic contents. In a semantic space of n dimensions, a sentence may encode $n-1$ contents.

Double entendres are conscious word games. However, there is a case to be made for the existence of cryptic *double entendres*. Consider the following anecdotes. My wife is a woman of color and is considerably younger than I am. When we make new acquaintances, amongst the first things that they notice is our difference in age and race. Of course, it would be socially unacceptable for them to overtly mention this, but they often react in a manner that can only be reasonably understood as an unconscious verbal response. On one occasion, immediately after having been introduced to us, a male academic suddenly remarked that his brother-in-law had recently adopted an African child, and that he (the brother-in-law) is really too old to do this. My colleague's narrative turned on the two themes of age difference and racial difference: the "too old" (and presumably Caucasian) brother-in-law adopting a Black child. Considering the social context, treating this as a cryptic *double entendre* is nearly inescapable. It sounds very much like he was informing me that he thought that I was too old to be married to this young Black woman. On another occasion, after a similar introduction, another male colleague spontaneously began to discuss the repertoire of the African-American jazz musician Louis JORDAN, and specifically mentioned one (and only one) of JORDAN's songs. He chose one of the more obscure pieces from this corpus entitled "That Chick's too Young to Fry", which is itself a double entendre describing a man contemplating a liaison with a girl who is, as the title implies, "too young" for him. Of course, these examples prove nothing, but they suggest a good deal. They illustrate a phenomenon that, I will argue, frequently occurs in situations of interpersonal conflict over biologically significant variables. This claim is easily corroborated by informal naturalistic observation and is to at least some degree testable using conventional empirical methodologies. Furthermore, it makes a good deal of sense in the context of the conception of the human mind entailed by the prevailing evolutionary account of the origins and dynamics of self-deception.

Self-Deception as a Cognitive Adaptation

I use the term "self-deception" to denote the exclusion of representational states from consciousness. At the most general level, there are two respects in which self-deception could be adaptive.

Representational states may be barred from consciousness in order to insulate certain psychological processes from conscious scrutiny (processes which achieve their ends most efficiently in the absence of conscious interference).

The processes in question would, if made conscious, disrupt other adaptive functions carried out by the conscious mind.

The definition given above differs from many others in that it does not describe self-deception as necessarily deliberate (in the vernacular sense of the word), motivated or indeed alterable. In fact, it allows for the possibility that self-deception is a function of our cognitive architecture, i.e., that certain types of information may be barred from consciousness because evolution has either dismantled or failed to construct the required neural pathways. The structural segregation of conscious from unconscious processes may have been naturally selected if this augmented the fitness of our ancestors.

The prevailing theory of the evolution and adaptive function of self-deception is based on the analysis of the dynamics of intraspecific deception. Deception is advantageous in social exchange because it allows one to "cheat" and reap a profit in what is ostensibly a reciprocal arrangement. The intentional deception of conspecifics (in contrast to forms of deception based on an organism's morphology or hard-wired behavioral routines) requires a high level of cognitive sophistication, including access to a Theory of Mind. It is not surprising to find that tactical deception has been observed in the higher primates and seems particularly prevalent amongst our cousins the chimpanzees (DE WAAL 1986; MILES 1986). Of course, *Homo sapiens* are superb deceivers: "We are primates who are experts in deceit, double-dealing, lying, cheating, conniving and concealing" (O'CONNELL 1998). There is no evidence that chimpanzees are capable of self-deception. As far as we know, *Homo sapiens* is the only species with this ability. It therefore seems probable that the capacity for self-deception arose somewhere along the hominid line.

TRIVERS' (1976, 1981, 1985a, 1985b, 1988, 1991, 2000) suggests that self-deception is a cognitive adaptation that enables one to deceive others more effectively than would otherwise be possible. According to his account, the aptitude for deception in a population works as a selection pressure for the evolution of cognitive mechanisms for deception-detection; those members of the population who possess sufficient psychological acuity and vigilance are less likely to be exploited, and therefore more likely to be reproductively successful, than their less insightful fellows. By the same token, the presence of newly evolved deception-detection neural technology becomes a selection pressure for the emergence and proliferation of more sophisticated deceptive abilities in a coevolutionary arms race spiraling through evolutionary time. With each iteration, the mechanisms of deception become more subtle and the deception-detection abilities more incisive. Our ancestors eventually became so excruciatingly conscious of their own deceptive strategies, of the likelihood of being found out by wary conspecifics with finely-honed perceptual skills, and of the possible reprisals by their community, that deception began to be accompanied by high levels of stress and anxiety. This, in turn, produced involuntary symptoms such as sweating, postural cues, changes in voice pitch, shifty eyes, etc. (EKMAN 1988, 1992), all of which conspired to betray the deceiver's true intentions to the expert mind-readers around them. According to TRIVERS' theory, the evolution of the ability to deceive oneself was nature's solution to this problem. By alienating themselves from their own duplicity, by rendering their own deceptiveness unconscious, by deceiving themselves about their own deceitful engagements, our ancestors were able pre-empt those telltale signs of stress and thus make it more difficult for others to apprehend them.

Biologists propose that the overriding function of self-deception is the more fluid deception of *others*. That is, hiding aspects of reality from the conscious mind also hides these aspects more deeply from others. An *unconscious* deceiver is not expected to show signs of the stress associated with consciously trying to perpetrate deception (TRIVERS 1988, p.vii).

Self-deception and the Evolution of Conscious Cognition

"Consciousness" is a vague and treacherous word. To use it is to invite confusion, but to avoid it entirely is scarcely possible. In this paper, I will use the more specific term "conscious cognition" instead of the more general term "consciousness" to refer an organism's awareness of it's own representational states (what philosophers call "intentional states'). By the same token, I will use "unconscious cognition" to refer to representational states of which their owner is unaware. There are, of course, many conscious and unconscious non-representational states that do not fall with in the purview of this definition.

Self-deception entails unconscious cognition, but unconscious cognition does not entail self-deception. There are many cognitive functions that remain unconscious because there is no need for them to become conscious. The unconscious processes described by cognitive science, such as MARR's (1982) account of visual processing, are typically of this nature.

TRIVERS (1991) expresses the view that unconscious cognition is the default position in nature, and that conscious cognition only enters the picture "under conditions in which it makes sense to invest extra energy and faculties to scrutinize something carefully" (p179). On this view, conscious cognition is a mode for dealing with the non-routine but vital situations with a high level of outcome variance. CLAXTON's (1994, 1998) position is consistent with this. He proposes that conscious cognition began as a response to threat or uncertainty.[1] Danger galvanizes attention: a threatened animal appears to be *hyperaware*. This is also true of a hunting animal stalking unpredictable prey. Perhaps our remote ancestors experienced episodes of conscious thought only at those moments when a twig snapping in the underbrush whispered a warning of a lurking hyena, or when devoting every ounce of attention to moving silently through the bush to spring upon an unsuspecting antelope. The vital ability to judge what predator or prey are likely to do, to anticipate their moment-to-moment moves, must have acted as a selection pressure for what psychologists call "mindreading" abilities: the ability to infer intentional states in others (BARRETT 1999; BOYER 2001). Having become established for this purpose, these skills could be redeployed for understanding the much more complex intentional states of our fellow hominids.

One would expect both deception and deception detection, and thus self-deception, to reach their apogee amongst social creatures living in large groups with intricate social networks rife with opportunities for cheating both by individuals and coalitions. Natural selection would favor the evolution of powerful and subtle social inference system amongst such creatures. According to the "MACHIAVELLIAN intelligence" hypothesis this is in large measure what drove the evolution of primate intelligence (HUMPHREY 1976; BYRNE/WHITEN 1988; WHITEN/BYRNE 1988, 1997; BYRNE 1993, 1994). The impressive cognitive powers of *Homo sapiens* may have been hammered out on the anvil of social conflict and deceit.

As our ancestors formed larger and larger social groups, the threat posed by large predators receded. It was at this point that we became our own worst enemy (as Jean-Paul SARTRE's character Garcin famously remarked "Hell is other people"). "Other human beings", writes Richard ALEXANDER, " became the principal hostile force of nature—the principal cause of failure to survive or reproduce" (ALEXANDER 1987, p78). One way of dealing with intraspecific predation was to form larger social groups (DUNBAR 1996). But this development also imposed costs. A linear increase in group size resulted in an exponential increase in social complexity, and therefore an exponential increase in the MACHIAVELLIAN dangers posed by one's own social group.

"In a group of five individuals I have to keep track of a set of four relationships between myself and the other group members, but I have to monitor six additional relationships involving the other four individuals. In a group of twenty, I have to keep track of nineteen relationships between myself and fellow group members, and 171 third-party relationships involving the other nineteen members of the group. While my relationships with everyone else has increased roughly fivefold with the fivefold increase in group size, the number of third-party relationships I have to keep track of has increased almost thirtyfold" (DUNBAR 1996, p65).

If expanding group size dramatically enlarged the opportunities for and dangers of exploitation, it seems at least plausible that the cognitive adaptations originally shaped to deal with predator–prey relations would be co-opted for dealing with the dangers of social life. The ability to anticipate hunting tactics of a hungry hyena were put to new use to monitor the MACHIAVELLIAN social tactics of conspecifics, and the cognitive adaptations responsible for predatory prowess were used for *social* predation: the deception and exploitation of one's own kind. This development would have been accompanied by a heightening of conscious social cognition. Whereas consciousness had been episodic and reserved for occasional predator–prey interactions, it now became chronic because of the ongoing vigilance required to navigate complex and potentially hazardous social systems.

Conscious cognition is focal, linear and has a very limited information-bearing capacity. In tasks involving complex stimuli, learning proceeds most effectively on an "implicit" or unconscious level, and conscious cognition is actually counterproductive (REBER 1976; REBER et al. 1980). The characteristics of conscious cognition are ideal for the highly focused tasks of monitoring predators and tracking prey in hunting situations, but spectacularly poorly suited for keeping tabs on the enormously complicated simultaneous interactions of hominid group life, and the intricate rules governing it. It is doubtful that conscious cognition would have been able to adequately manage the new cognitive burden placed upon it. I suggest that the problem of conscious cognitive overload was solved by outsourcing a good deal of social cognition to the massively parallel unconscious cognitive apparatus.[2]

In addition to solving the problem of conscious cognitive overload, this division of mental labor undercut the tendency for deceivers to betray themselves by showing signs of stress, and thereby facilitated interpersonal deceit. On this analysis, unconscious social cognition must include at least two interlocking mental functions:

(a) MACHIAVELLIAN social planning, including deception, manipulation, as described by TRIVERS and others who have embraced his theory of self-deception.

(b) The capacity to monitor the behavior of those with whom one interacts. This feature is implied by TRIVERS' theory, because effective deception requires perceptual feedback. Success in tactical deception would be very unlikely if deceivers were unable to carefully track the moment-to-moment responses of their victims, and to adjust their social choreography accordingly. Given the need to conceal deceptive engagements from consciousness, this inferential process must be assumed to occur unconsciously, to be biased in favor of its owner's interests (because its function is to enhance its owner's fitness), and to be somewhat "paranoid" with respect to the intentions of others (i.e., to be highly sensitive to possible deceptive and exploitative ploys by others).

This line of investigation suggests that *Homo sapiens* possess an unconscious organ specifically devoted to tracking complex MACHIAVELLIAN social exchanges and making rapid decisions about the implications of other peoples behavior with regard to conflicts and confluences of interest concerning biologically significant variables. This unconscious mental system should be particularly sensitive to the *grammar of social exchange*: the implicit norms governing social intercourse which define the parameters of costs to and benefits for agents. We would expect the main input for this system to be non-verbal information. Words are ideally suited for deception because of their representational power and their biological inexpensiveness (TRIVERS 1985b). Whereas speech entails minimal investment of resources, actions are relatively costly, making them far more likely than words to reveal an agent's true interests.

Where Is the Data?

The existence of unconscious mental processes has only recently been widely accepted by psychologists, although their nature remains in dispute. Contemporary research into the unconscious is mainly focused on the study of non-affective cognitive processes in laboratory settings. It would be most unlikely that such research would throw light on the processes described in the present paper, which are *affectively charged social processes* and are probably best observed in naturalistic or quasi-naturalistic settings. The relevant data should be all around us, saturating our social lives. Be that as it may, there are very few situations in which observations of such interactions have been systematically recorded.

One setting in which affectively charged conversations regularly occur, and are recorded at least to some degree, is one-to-one psychoanalysis or insight-orientated psychotherapy. Here we find an emotionally charged encounter with another person in which the rules of conduct are ambiguous, violating many social norms and expectations. It is also a verbally permissive setting, in which the analysand is encouraged to say everything and anything that comes to mind. In light of these features, it would be remarkable if protocols of psychoanalytic sessions did not reveal evidence of unconscious social cognition, at least on the part of analysands. Of course, using psychoanalytic data in this way does not entail an endorsement of psychoanalytic clinical or theoretical claims.

"Psycho-analysis has shown us that everyone possesses in his unconscious mental activity an apparatus which enables him to interpret other people's reactions, that is, to undo the distortions which other people have imposed upon the expression of their feelings" (FREUD 1958, p139).

Notwithstanding this demurer, it is rather interesting to note that FREUD occasionally described a process that sounds suspiciously like unconscious social cognition. "I have had good reason for asserting", he wrote, "that everyone possesses in his own unconscious an instrument with which he can interpret the utterances of the unconscious in other people" (FREUD 1953, p320).

There are scattered references to the same idea in the post-FREUDIAN literature, but nothing like a systematic discussion prior to the work of LANGS, who argued that incisive unconscious perceptions are unconsciously transformed into superficially unrelated narrative sequences (such as the story that my colleague told about his brother-in-law) which nonetheless express affectively significant information (LANGS 1978, 1979, 1980, 1981, 1992; SMITH 1999).

A second rich source of data on affectively charged interactions comes from work on T-groups, small groups consisting of a trainer and a number of participants who are strangers to one another, are used for the study of group processes and for self-exploration. The T-group setting resembles the psychoanalytic setting insofar as the rules for social engagement are ambiguous, and the participants are encouraged to speak freely and engage in self-disclosure. HASKELL, a cognitive psychologist, has discovered patterns of communication in T-group data that suggest the operation of unconscious social cognitive processes (HASKELL 1978, 1987, 1988, 1990, 1991, 1999a, 1999b, 2001). He has developed a complex model of how an array of unconscious cognitive and psycholinguistic operations are deployed to select and manipulate narrative images which, when introduced into ordinary conversations, carry unconscious meaning. HASKELL refers to this process as "sub-literal" communication, a term which I will adopt.

Whatever our points of disagreement in other respects, the work of HASKELL, LANGS and I converges on five crucial points.

1. Sub-literal communications are vehicles for messages of a kind that are not normally tolerated in polite conversation. To speak sub-literally is to speak the unspeakable. The *meaning* of sub-literal communications violates social taboos, but their *form* disguises this fact.

2. Sub-literal communication is unconscious. Even in very transparent examples, the speaker is liable to deny that they had any such thing (consciously) in mind.

3. Sub-literal communication is lawful. The unconscious selection of narrative imagery covaries with events occurring in the immediate conversational setting.

4. Sub-literal communication is responsive to the fundamental structure of the conversational setting and to biologically significant variables active within that setting (e.g., dominance, cheating, resources, sex, coalitions, etc.).

5. Sub-literal information is expressed through the medium of spontaneous narratives, usually social narratives describing the activities, attributes and interactions of other people.

There is evidence from cognitive science suggesting that narrative and non-narrative modes of communication are based on different forms of cognition (e.g., BRUNER 1986; BUCCI 1997). According to DONALD (1991) and DUNBAR (1996) narrative mode is probably the most ancient mode of verbal communication. Social narration (gossip) is perhaps the most widespread form of narrative activity, occupying 60 to 70 percent of conversation time (DUNBAR 1996; EMLER 1992). Evolutionary psychologists have conjectured that our irrepressible proclivity for gossip may have had evolved as a mechanism to spread information about cheating within ancestral groups (COSMIDES/TOOBY 1987; EMLER 1992; ENQUIST/LEIMAR 1993). This is consistent with the hypothesis that sub-literal communication is an output from a mental module the proper function of which is to track and analyze MACHIAVELLIAN social interactions.

Illustration

The following example from a psychotherapeutic session was specifically selected to illustrate the phenomena described in the present paper, although it is typical of the richly patterned sub-literal communications that can be observed in such settings.

The client is a fourteen year-old girl, whom I will call "Jane", who is in once-weekly therapy with a female psychotherapist in London, England. The girl's mother is also being seen by another psychotherapist in the same clinic. During the week prior to the session that I will present, the Jane's mother telephoned the therapist to discuss her daughter's difficulties. This is their second session.

Jane: Jessica is still playing up. She has now got her mother to ring up the school so that the teacher could speak to Julia and me. She want's us to be nice to Jessica. I don't think she should have done this behind our backs [pause]. No wonder we can't be best friends. I thought I could trust her.

Jane begins with a social narrative involving a conspiratorial telephone conversation that went on behind her back. She emphasizes that this has led to a loss of trust. This narrative would appear to allude sub-literally to the clandestine conversation between the therapist and her mother, and it's ramifications for her relationship with the therapist. In biological perspective, Jane is confronted with a potentially hostile coalition between the therapist and her mother, in which the mother's interests may prevail over her own.

Therapist: You sound very disappointed with Jessica, as though she betrayed you.

Jane: I just think she is very weak. She should have told me how she felt rather than doing this behind my back. Now she has involved her mother and the whole school in this. [silence] I am angry with her. Last Saturday after I had been to see you, Julia and I went to the cinema. We had made arrangements for this in class and we knew Jessica had overheard us. And who should be there but Jessica! She pretended not to see us and when we asked what she was doing there it was all an accident, that she didn't know we were going to be there. But I knew she knew. I knew I was right but she would never admit to this.

The therapist does not allude to the telephone conversation in her intervention. In light of the hypotheses presented here, one would expect Jane to unconsciously view this as a MACHIAVELLIAN maneuver to deceive her about her coalition with the mother. One would also expect this to be expressed sub-literally in Jane's narrative discourse following the intervention through stories about interpersonal deceit. These expectations are amply confirmed. Jane appears to use Jessica to represent the therapist, stating that Jessica should have spoken to her directly rather than going behind her back. Jessica lied, and although Jane knew what was going on, Jessica "refused to admit" the truth.

Therapist: What seems to have upset you about this incident is that your perception of what had happened was not confirmed. Instead Jessica pretended it was all a coincidence but you knew she had overheard you. This reminds me of how you described feeling in relation to your parents. They both give you different versions of events. Even though you tell me you know what was happening at the time, you sometimes start doubting.

Jane: I do get confused but I remember what happened. I just wish that my mother would accept that although I was young at the time of their divorce I was still aware of what was happening. [silence] She always uses the excuse that because I was young I don't know what I'm talking about…that I can't remember correctly. Then she goes on and on and tells me her version. But I remember she was drunk, and that she is the one who can't remember.

Once again the therapist refrains from mentioning her telephone conversation, and attempts to divert Jane's attention onto her relationship with her parents. Jane responds with a narrative about how her mother tries to get her own faulty version of events to prevail over Jane's true recollections. This appears to be a sub-literal commentary on the MA-

CHIAVELLIAN implications of the therapist's intervention, with Jane's mother now representing the therapist, whose omission of any mention of the telephone conversation is unconsciously interpreted by Jane as an attempt to bamboozle her. The client's remark that "she goes on and on and tells me her version" may be a representation of the therapist's lengthy and, from Jane's perspective, deceitful intervention.

Therapist: How does it feel when she tells you such things?

Jane: I lose confidence in myself and I get angry. I often go and hide in the toilet just to get away from her. When I was little I used to go up to my room and lie on my bed talking to my teddy bears. I liked talking to them because I knew I could trust them. I could talk to them and they would not tell anyone what I had said because they could not speak. I could cry with them.

The therapist, continuing in much the same vein, elicits a narrative in which Jane expresses her desire for a relationship with someone who will not divulge information to a third party.

Therapist: You started off your session today expressing your disappointment and anger with Jessica. You described her as being weak. Instead of telling you up front what was wrong she got her mother to call the school. By doing this she brought other people into something which concerned just you and her. Also, last week when you met at the cinema she pretended that your meeting was a chance meeting even though you knew that she had overheard your arrangements with Julia. You felt she was unwilling to admit to this and it seems important to you that people are honest. You also described how you could only talk to your Teddy bears as they were the only ones with whom you could have a unique, safe relationship without bringing anyone else in. I think you might feel betrayed by me because not only have I had some contact with your mother when she called me but I have not mentioned this to you. You seem to be stressing that your mother, or anyone else, should not be involved in this relationship. It should be like the relationship you had with your Teddy bears or else you feel betrayed.

The therapist finally comes clean about the telephone conversation, and acknowledges that her client feels that this was antagonistic to her best interests. At this point, we would expect the content of Jane's narratives to change. The theme of deception should disappear, because the therapist is no longer deceiving her, and should be replaced with narrative imagery expressing confluences rather than conflicts of interest.

Jane: [pause] After my first session, Mum asked me how it went. I just said I liked it. I needed someone to talk to but there are some things I just can't tell her. Dad also asked me why I came to see you. [silence] I can talk to Julia. She really understands. We really get on. She really is my best friend. It's not that we don't argue but she seems to know when I'm upset. When I go all quiet she will eventually come up to me and ask what is wrong so that we can talk about it.

Therapist: It's time to finish now.

Testing the Theory: The Science of Gossip

The examples that I have provided were specifically chosen for their value as compelling illustrations of sub-literal communication. As such, they prove nothing. I could have provided many hundreds of similar examples and have still been vulnerable to the charge that these were chosen specifically because they conform to my theoretical beliefs and are not representative of the structure spontaneous social narrative. This objection is entirely reasonable.

The only way to neutralize the charge of selection bias is to make use of an appropriately regimented empirical methodology capable of demonstrating the systematic covariation of narrative themes with biologically significant variables in the conversational setting. There are two broad strategies by means of which this can be accomplished:

1. Using existing protocols of psychotherapeutic or T-group sessions, use the theory to predict the general content of unfolding narrative imagery, or to retrodict significant variables from the specific narrative imagery unfolding in the session.

2. Design an experimental situation in which affectively significant biological variables are manipulated by the investigator. The content of subjects' narrative activity (which can be easily accessed by, for example, asking them to write a story about people in their neighborhood), as the dependent variable, should covary with the independent variable in specific and predictable ways.

Author's address

David L. Smith, University of New England, 11 Hills Beach Road, Biddeford, ME 04005, USA. Email: dsmith@une.edu

Acknowledgments

I would like to thank my wife Subrena SMITH for her observations on the adaptive functions of unconscious social cognition and the significance of unconscious perceptual feedback loops, as well as for her comments on an earlier version of this paper. I would also like to thank Dr. Robert HASKELL for his helpful suggestions.

Notes

1 Both TRIVERS and CLAXTON use the more general terms "consciousness" and "unconsciousness". However, I interpret them to be referring to cognitive states.

2 A version of this thesis was initially suggested by LANGS (1995, 1996).

References

Alexander, R. D. (1987) The Biology of Moral Systems. Aldine de Gruyter: New York.

Barrett, H. C. (1999) Human cognitive adaptations to predators and prey. Doctoral dissertation. University of California: Santa Barbara.

Boyer, P. (2001) Religion explained: The evolutionary origins of religious thought. Basic Books: New York.

Bruner, J. (1986) Actual minds, possible worlds. Harvard University Press: Cambridge, MA.

Bucci, W. (1997) Psychoanalysis and cognitive science: A multiple code theory. Guilford Press: New York.

Byrne, R. W. (1993) Do larger brains mean greater intelligence? Behavioral and Brain Sciences 16(4):696–697.

Byrne, R. W. (1994) The evolution of intelligence In: Slater, P. J. B./Halliday, T. R. (eds) The Evolution of Behaviour. Cambridge University Press: Cambridge (England), pp. 225–265.

Byrne, R. W./Whiten, A. (1988) Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans. Oxford University Press: New York NY.

Claxton, G. (1994) Noises from the darkroom: The science and mystery of the mind. Aquarian: London.

Claxton, G. (1998) Hare brain tortoise mind. Fourth Estate: London.

Cosmides, L./Tooby, J. (1987) Evolutionary psychology and the generation of culture, Part II. Ethology and Sociobiology 10:51–97.

de Waal, F. (1986) Deception in the natural communication of chimpanzees. In: Mitchell, R. W./Thompson, N. S. (eds) Deception: Perspectives on human and nonhuman deceit. State University of New York Press: Albany, pp. 22–244.

Donald, M. (1991) Origins of the modern mind. Harvard University Press: Cambridge MA.

Dunbar, R. (1996) Grooming, gossip and the evolution of language. Faber and Faber: London.

Ekman, P. (1988) Self-deception and detection of misinformation. In: Lockard, J. S./Paulhus, D. L. (eds) Self-Deception: An Adaptive Mechanism. Prentice Hall: Englewood Cliffs NJ, pp. 229–250.

Ekman, P. (1992) Telling lies: Clues to deceit in the marketplace, politics and marriage. W.W. Norton & Co: New York.

Emler, N. (1992) The truth about gossip. Social Psychology Newsletter 27:23–37.

Enquist, M./Leimar, O. (1993) The evolution of cooperation in mobile organisms. Animal Behaviour 45:747–757.

Freud, S. (1958) The disposition to obsessional neurosis. In: Strachey, J. et al. (eds) The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume 12. Hogart PressLondon, pp. 317–326. Originally published in 1913.

Freud, S. (1953) Totem and taboo. In: Strachey, J. et al. (eds) The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume 13. Hogart Press: London, pp. 1–161. Originally published in 1913.

Haskell, R. E. (1978) An analogic model of small group behavior. International Journal of Group Psychotherapy 28:27–54.

Haskell, R. E. (1987) Social cognition and the non-conscious expression of racial ideology. Imagination, Cognition and Personality 6(1):75–97.

Haskell, R. E. (1988) Small group "fantasy theme" analysis: Anthropology and psychology: A comparative study of a psychosocial structure of a ritual ceremony. Journal of Psychohistory 16:61–78.

Haskell, R. E. (1990) Cognitive operations and non-conscious processing in dream and waking reports. Imagination, Cognition and Personality,10:65–84.

Haskell, R. E. (1991) An analogical methodology for the analysis and validation of anomalous cognitive and linguistic operations in small group (fantasy theme) reports. Small Group Research 22:443–474.

Haskell, R. E. (1999a) Between the lines: Unconscious meaning in everyday conversation. Plenum/Insight: New York.

Haskell, R. E. (1999b) Unconscious communication: Communicative psychoanalysis and sub-literal cognition. Journal of the American Academy of Psychoanalysis 27(3):471–502.

Haskell, R. E. (2001) Deep listening: Hidden meanings in everyday conversation. Perseus: Cambridge MA.

Humphrey, N. (1976) The social function of intellect. In: Bateson, P. P. G./Hinde, R. A. (eds) Growing points in ethology. Cambridge University Press: Cambridge (England), pp. 303–317.

Langs, R. J. (1978) The listening process. Jason Aronson: New York.

Langs, R. J. (1979) The therapeutic environment. Jason Aronson: New York.

Langs, R. J. (1980) Interactions: The realm of transference and countertransference. Jason Aronson: New York.

Langs, R. J. (1981) Resistances and interventions: The nature of psychotherapeutic work. Jason Aronson: New York.

Langs, R. J. (1992) Science, systems and psychoanalysis. Karnac: London.

Langs, R. J. (1995) Clinical practice and the architecture of the mind. Karnac: London.

Langs, R. J. (1996) The evolution of the emotion processing mind. Karnac: London.

Marr, D. (1982) Vision. Freeman: New York.

Miles, H. L. (1986) How can I tell a Lie? Apes, language and the problem of deception. In: Mitchell, R. W./Thompson, N. S. (eds) Deception: Perspectives on human and nonhuman deceit. State University of New York Press: Albany, pp. 254–266.

O'Connell, S. (1998) Mindreading: How we learn to love and lie. Arrow: London.

Reber, A. S. (1976) Implicit learning of synthetic languages: The role of instructional set. Journal of Experimental Psychology: Human Learning and Memory 2:88–94.

Reber, A. S./Kassin, S. M./Lewis, S./Cantor, G. W. (1980) On the relationship between implicit learning of a complex rule structure. Journal of Experimental Psychology: Human Learning and Memory 6:492–502.

Smith, D. L. (1999) Approaching psychoanalysis: An introductory course. Karnac: London.

Trivers, R. L. (1976) Foreword. In: Dawkins, R., The selfish gene. Oxford University Press: Oxford.

Trivers, R. L. (1981) Sociobiology and politics. In: White, E. (ed) Sociobiology and human politics. Lexington Books: Lexington MA.

Trivers, R. L. (1985a) Social evolution. Menlo Park CA: Benjamin/Cummings.

Trivers, R. L. (1885b) Interview. Omni, July:111.

Trivers, R. L. (1988) Introduction. In: Lockhard, J. S./Paulhus, D. L. (eds) Self-deception: An adaptive mechanism? Englewood Cliffs NJ: Prentice Hall, pp. vii–ix.

Trivers, R. L. (1991) Deceit and self-deception: The relationship between communication and consciousness. In: Robinson, M. H./Tiger, L. (eds) Man and beast revisited. Smithsonian Institution Press: Washington, pp. 175–191.

Trivers, D. L. (2000) The elements of a scientific theory of self-deception. Annals of the New York Academy of Sciences 907:114–131.

Whiten, A./Byrne, R. W. (1988) Tactical deception in primates. Behavioral and Brain Sciences 11:233–273.

Whiten, A./Byrne, R. W. (eds) (1997) Machiavellian intelligence II: Extensions and applications. Cambridge University Press: Cambridge.

H. John Caulfield

An Account of Self-Consciousness

Its Evolution and Extension to Human Artifacts

Abstract

To each reader, his self is a matter of instinctive primary concern. Without attending much to the philosophers of self, I have sought to ask if a self and self-consciousness might evolve simply and naturally in some conscious creatures. I suggest a simple path for such evolution and show why, once it evolves, self consciousness is very likely to be favored by evolution. With that model of what self consciousness is, I show how to provide it for complex human artifacts, why it would be desirable to do so, and how we humans should interact with such systems so they carry out our goals.

Key words

Self, consciousness, evolution.

Introduction: Meta Cognitive Sciences as Applied here to Self-Consciousness

Physics is underdetermined by reality. That is, there are multiple physical theories that account equally well for observations. That accounts for the large number of interpretations of quantum mechanics, for instance. To choose among the theories, physicists must rely not on measurements (physics) but on criteria about physical theories—metaphysics. The only widely accepted metaphysical principle is OCCAM's razor, and much dispute arises as to how to apply that—metaphysical issues. Endless regress is unavoidable.

Likewise, in choosing among theories of the nature and origin of self and self consciousness, we must be guided by meta cognitive sciences criteria. I offer mine here. They can only be attacked or defended on meta cognitive sciences grounds. I do not propose to visit such a defense here. I offer my criteria only to show my biases explicitly, and hope to have this work judged within the context of those criteria.

1. No magic. I reject accounts that have the "miracle" of self either appear magically from outside conventional physical theory either from the beginning (panpsychism) or as something that happens only at some late period of evolution. No insight from revelation (mine or someone else's) is allowed. Such revelations may or may not have value in other contexts, but they are excluded here. Lovely examples of magic include ECCLES (1991) and PENROSE (1999). Both seem to seek something outside of current science as an explanation. See also (CHALMERS 1996) and my review of it (CAULFIELD 2002).

2. Quantum mechanics is to be invoked if and only if simpler classical explanations fail. The reasoning of quantum mind proponents (PENROSE 1987; STAPP 1998) is (implicitly) that mind and quantum mechanics are both mysterious, so one must account for the other. I see no need for dragging quantum mechanics into something explainable on a simpler level.

3. Every aspect of animal life must have resulted from evolution; accordingly I will accept theories of self consciousness that could arise with minimum changes from the prior state (consciousness) and have an evident survival value to its possessor. This seems to me to exclude epiphenomenology. There is a whole journal dedicated to this aspect of the discussion—EVOLUTION & COGNITION.

A theory of self-consciousness is defective if it does not provide an understanding on how to produce self-conscious artifacts—robots (R2D2), ships (HAL), and so forth. A consequence of this is that I seek an understanding that is somewhat independent of the neurobiology of animal self consciousness. It must account for the evolution of animal self consciousness, but the account sought must be broad enough to encompass nonbiological entities as well. FREEMAN/WATTS (1941) among others have discussed the neurobiology of self consciousness admirably. This year, an important new book on this subject was published by LEDOUX (2002). My goal is to applaud not critique such efforts, while aiming at understanding what the phenomenon of self-consciousness is, so (among other things) we will be able to build self-conscious artifacts. Likewise, BAARS (1996) has a nice analysis of self-consciousness with which I have no substantive quarrel. I caution, however, that BAARS is among many cognitive scientists who seem to use consciousness to mean what I call self-consciousness and, therefore lack an explicit theory of primitive consciousness.

4. Contrary to the philosophers who denigrate "folk psychology", (e.g., CHURCHLAND 1989) I will accept an account of self-consciousness only if it can account for what ordinary human beings experience as self-consciousness. Explanations that fail this criterion are to be considered explanations of something else—not explanations of self consciousness.

Background and Overview

It is much easier to provide a background on theories of self consciousness that do not meet those criteria than to discuss those that do. So far as I know, no theory meeting those criteria has ever been offered. Indeed that is the purpose of this paper.

I think the closest work is that by STRAWSON (1997). This is a very good philosophical paper that reaches almost exactly the same conclusion I will reach. In particular, he emphasizes the experience we humans have of a self that is somewhat independent of time. I call such a self "aternal". He calls it "diachronic". Indeed our conclusions are consistent with Western philosophy (quotations from STRAWSON 1997).

- "The soul, so far as we can conceive it, is nothing but a system or train of different perceptions" (HUME 1739).
- "What was I before I came to self-consciousness? … I did not exist at all, for I was not an I. The I exists only insofar as it is conscious of itself… The self posits itself, and by virtue of this mere self-assertion it exists" (FICHTE 1794–1795).
- "The 'Self'…, when carefully examined, is found to consist mainly of… peculiar motions in the head or between the head and throat" (JAMES 1890).

It is also the essence of much eastern religion and western metaphysics. What the Buddha called "enlightenment" is the realization that the aternal self is a "self created" fiction—*maya*. So, we do not have to be concerned about our future self. There is no such creature. Jesus of Nazareth is reported to have said much the same things: "Take no thought for the morrow, for the morrow shall take thought for things of itself", "He who would save his life will lose it", and many other such statements.

If my conclusions are as ancient as 2500 years, what do I hope to add? I hope to show that such a self consciousness arises within the quite expectable (*a posteriori*) course of evolution, and that it is quite functional, understandable, and extendable. At the same time, the fact that evolution so strongly favors self consciousness allows us to understand why even Buddhas (enlightened individuals who know better) can never fully overcome "selfishness". The latter assertion corresponds to the ancient Christian doctrine of "original sin".

Other modern philosophers have reached the somewhat the same conclusions. DENNETT (1991), for instance, calls the self the "Center of Narrative Gravity"—a statement so clear as to need no elaboration. A significant recent compendium of work in this field was published recently (GALLAGHER/SHEAR 2000) and provides extensive background that I feel no need to repeat here. I can recommend the book by MACPHAIL (1998) as an excellent summary of work on this field, albeit one aimed primarily at nonhuman animal consciousness. Another nice book that reaches somewhat similar conclusions was also published recently by BERMUDEZ (2001). There are aspects of BERMUDEZ's work with which I might disagree, but its breath and fairness warrant its inclusion here. Much of this can be found in discussions on his work with POWER (2001).

Again, my aim is not to reach a philosophical conclusion, so philosophers how ever profound are not necessarily helpful to either my writing or the reader's understanding of what I write. My aim is to give an account of self-consciousness that satisfies the criteria just listed.

Every reader of this article feels that he is a self—an entity that exists and persists through time and has a body and a mind. He is concerned about his current and even his past and future. This complex of observations is what I seek to discuss in terms of both its evolution and its possible extension to our artifacts—robots, computers, and so forth.

The hypothesis I defend is that a creature that has self consciousness has a profound evolutionary advantage over one that does not. Without self consciousness, our lives would appear to be real only in the extended now. A creature that views itself as being a self that is somehow not only now but also then and would tend to the interests of that self. In particular, it would do things now for the future good of that self. Therefore, it would have better long-term survival prospects than a creature that lives only in the instant. The model of primitive consciousness I discussed earlier might easily allow self-consciousness to arise. And when it does, evolution is likely to preserve and enhance it. It is an important side note to indicate that many animals have evolved instincts that have a positive impact on their futures. A good example might new those mammals that build underground dwellings with multiple entrances/exits. Almost certainly, they are not consciously planning escapes from future predators. They just "felt like" making such a dwelling. I am concerned with conscious planning for the future—an entirely different thing. Humans use instinct as well as conscious planning.

A creature with both consciousness and self consciousness may still lack the rich mental and emotional life of a mature human. Humans have added dimensions of coevolved language and culture that is not a necessary feature of self consciousness. Indeed, I will argue below that self consciousness extends at least through apes and cetaceans.

Understanding self consciousness in this way allows us to make sense of philosophical and religious insights from all human cultures. I note this, because a theory of self-consciousness that failed in that sense might well be held to be inadequate to explain what humans experience. It also allows us to make sense of the mirror image recognition test widely used to test for self consciousness.

The state of Alabama pastes a sticker in biology textbooks to remind students that no-one knows that evolution occurred, because no one was there to witness it. That is a problem for anyone who would seek to understand how and why something evolved. The best we can do is to show that the feature in question is rather easily reachable from a stage of evolution immediately preceding the advent of the feature to be explained and to show that if that feature did arise, it would give its possessors a survival advantage. All of that proves nothing, just as Alabama legislators observe. On the other hand, an hypothesis that can do those things is certainly more plausible than one that cannot. I seek to defend such a plausible hypothesis.

My argument has two parts. First, self consciousness is an extension of primitive consciousness (CAULFIELD et al. 2001). It does not require the sudden magical appearance of something totally new. It is something that is likely to occur given sufficient time in conscious animals. Second, once it does appear, it will thrive, because it will make its possessors significantly more likely to produce and nurture to sexual maturity.

What I Mean by Consciousness and Self-Consciousness

We are all self conscious creatures. We understand our lives in a rich subjective way. It is very hard to imagine what it would be like to be conscious without self consciousness (what I call primitive consciousness) or even self conscious without the human cultural and language features that have coevolved with us. This requires an unusual act of imagination. The burden of this paper is to persuade the reader that what it describes is self consciousness at its most primitive level.

In a prior paper (CAULFIELD et al. 2001), we argued that a system that forms models of its body and its world and makes decisions based on those models in such a way that it tends to benefit the model body in its model world as judged by its limbic system can be said to be conscious of its body and world. This is what I called primitive consciousness.

In this paper, I argue that if the body is modeled as something with indefinite temporal extension into both past and present, then the good of that aternal body becomes the concern. That aternal creature (or model of a creature) that occupies a body and experiences a life is a primitive self. Such a creature is self conscious.

The Evolution of Self-Consciousness from Consciousness

According to this understanding, the critical element is the time aspect of the body model. In primitive consciousness, the body is modeled as something that exists in a slightly extended now. Creatures with only primitive consciousness are concerned about their life at the moment. They are untroubled about either past or future. The time aspect of the body model of a self conscious creature expands dramatically and indefinitely. They become selves with pasts and futures.

Thus the only step required to move from primitive consciousness to primitive self consciousness is to broaden the time span that will be considered in decision making.

How that time extension actually occurred is, of course unknown. But, here is what seems to me a plausible route. The route is through episodic memory. If a conscious creature remembers a prior event, it will tend to "identify" with the body that experienced and remembered the event. That identification creates the model self. Before, only a body now was of concern. Now, a self outside of time is of concern. The creature now and in the past is the same. The self persists through time.

A self model that persists through time automatically persists into the future as well. More importantly, there is little the creature can do about its past, but there is something it can do about the future self. It can take action now that will impact the self in the future. This requires a modified limbic system. Even in humans, we see the now-vs.-then decisions made very differently by different people. Some spend their lives preparing to live their lives in a future that never comes. Others live only for the moment and are generally called irresponsible. Most seek some sort of trade-off between now and then.

Thus, self consciousness can be seen as a modification of structure required for consciousness. It may well have arisen with episodic memory, as the latter almost certainly entails a point of view. I remember the event as I experienced it in some now. Thus somehow, the nows are the same. The I who is writing this paper is the same as the I who experienced a sixth birthday party many decades ago. With episodic memory comes a self that is no longer tied to the moment. It seems a small step to envision that self as having a future as well.

In humans, the aternal self has been built into our culture and language as well as our brains. It seems obvious to humans that we are aternal selves. Some but not all religions extend the time of a self indefinitely though reincarnation or heaven and hell. They take the aternal to eternal. Humans find it difficult to believe that the aternal self is a model made by a being now as opposed to an objective reality.

Selection

This should be the least controversial section of the paper. Evolution selects for traits that aid survival of the species the individual gene bearer and his offspring who are likely to carry its genes. Conscious planning for future well being is certainly likely to provide enough of an edge to be selected rather strongly. A creature that seeks the well being of its enduring self as opposed to the well being of its body now (as a merely conscious creature does) will fare better in the world. It will plan for its future.

Long before the evolution of self-consciousness, nature selected for animal traits that prepared instinctively for their futures and the futures of their offspring. Birds feel compelled to build nests. Squirrels feel compelled to hide nuts. Human mothers feel compelled to adore a smiling baby just as the baby is compelled to smile.

With the self-conscious creature, a new phenomenon arises—planning within the lifetime of an individual for the specific contingencies of his anticipated future life. Saving accounts are a wonderful proof of the self-consciousness of humans. Insurance policies are another proof.

Recognizing Self-Consciousness Experimentally

There is a standard test for self consciousness. If the baby human or ape or dolphin can recognize itself in a mirror, it is held to be self conscious (MITCHELL 1993; ANDERSON 1984; BUTTERWORTH 1995). Until writing this paper, I was always puzzled by this. Clearly recognizing you self in a mirror is somewhat related to self consciousness in that it involves identifying something that is not my body with something that is. Now, I can make more sense of it. In order to learn about mirror images, a creature must carry out accidental or deliberate experiments—touching it, touching its body, making faces, etc. In order to assimilate the information gained in such an experiment, the creature must have episodic memory. And, it must identify its current being with the subject of those past experiences. In short, it must be self conscious.

Self-Consciousness in Human Culture

If this explanation of self consciousness is correct, it must be consistent with what we humans experience as our self consciousness. It cannot offer a complete explanation, because humans have more than primitive self consciousness. Our culture and language teach us how to be self conscious beings.

I argue that the philosophies and religions of mankind today and for as far back as we can infer something about them have some commonalities so far as the human self is concerned. We are all taught that

- Each one of us *is* an individual self that has a past and a future
- We should be concerned not only about our selves but also about other selves. Indeed, individual human selves are the most important things in the universe. Treating others that way is the essence of morality.

Religions teach us different things about the self. The great Western religions—Judaism, Christianity, and Islam—all teach that there is a God who has created human selves (some say "souls") and is greatly concerned for their well being. The aternal self either is or can become an eternal self, dwelling in comfort (heaven) or discomfort (hell) depending on their performance, faith, etc. The great Eastern religions—Hinduism, Buddhism—teach that our sense of being a self is somehow illusory. Our task is to overcome that illusion and realize that we are in fact one with the universe. This summary is greatly simplified, of course. Western mystics have all taught that we are one with God but must learn to realize it. The Buddha taught that the belief that we are a self is the source of all of our fear and pain. If we understand that we are episodes in the eternal drama whose reality is only in the now, there would be no future me to worry about. This is what he called enlightenment, and one who has this realization is called a Buddha. But the mistaken belief (called *maya*) that we are separate selves is very hard to eliminate. Jesus of Nazareth had much to say on this subject. For instance, he advised that we not worry about what will happen tomorrow—the same advice offered by the Buddha.

Christianity asserts that "selfishness" is the root problem and also asserts that it cannot be overcome. It is built in—"original sin". Maya and original sin cannot be overcome, because it IS self consciousness. Evolution, language, and culture have all built it into us.

Self-Conscious Artifacts

This view of self consciousness shows how self consciousness can be built into robots or other systems with sensor, effectors, and computational systems in interaction with a world. This is, in fact, not hard if we start with a conscious artifact. Such an artifact has a world model, a body model, and a limbic system that it uses to seek the good of the model body in its model world.

Here is what we must do to make such a system self conscious.

Give it episodic memory

This is something some current neural networks can do without risking clutter problems. This is not he place to review such methods.

Provide each memory with a single point of view

That is these memories must be body centered and the body in all cases must be identified as one. That one, then, is no longer tied down in time. The same body can be modeled in past or future conditions.

Project the body's future state

This is the essence of predictive control. Brains are very good at it. They use neural networks that learn to associate future states with current states. These are very fast even with the very slow neurons they must work with.

Provide each future state's subject the same identity as the current and past states.

This is easy in computers, because we can simply define that identity in software.

Make the well being of the body extended in time the concern of the limbic system.

This leaves open the means for weighting present vs. future good in making decisions. As noted earlier, humans have adopted many different approaches to that problem. There are two extremes. The graduate student who sacrifices his social life, immediate income, and sleep to attain a degree that will allow "him" greater wealth and satisfaction years from now is one example of a limbic system that emphasizes long term good over short term reward. The other extreme might be represented by former President CLINTON who valued immediate pleasure so much as to be willing to risk his long term well being to accomplish it. Most humans operate between those extremes most of the time. Building in those trade-offs into an artifact is essential for its self-consciousness.

Such an artifact would have and be a self effectively outside of time. It would be self-conscious.

Conclusions

Our selves are our own creation. We cannot choose to be selfless, despite religious exhortations not to be selfish. We are "selfish" for the same reason we enjoy good food, attractive potential sexual partners, and young babies. Evolution built it into us. One may be a Buddha intellectually, but concern for the well being of one's aternal self always arises, because it must. That is *maya*. That is original sin. Opposing it is the *jihad* of the heart.

Author's address

H. John Caulfield, Fisk University, 1000 17th Ave., N., Nashville, TN 37208, USA. Email: hjc@fisk.edu

In some real sense, the future me is meaningful. It will be a reincarnation of the immediately preceding me, based on its memories, limbic system, instincts, dispositions, and so forth. It is entirely reasonable for me to treat you as an ongoing even aternal soul occupying your body. You behave that way.

My favorite Zen saying captures this situation accurately.

"You are not who you think you are.
Nor, are you other".

References

Anderson, J. R. (1984) The development of self-recognition: A review. Developmental Psychobiology 17:35–49.

Baars, B. J. (1996) Understanding subjectivity: Global workspace theory and the resurrection of the observing self. Journal of Consciousness Studies 3:211–216.

Bermudez, J. L. (2001) Bodily self-awareness and the will: Reply to Power. Mind and Machines 11:139–142.

Butterworth, G. (1995) The self as an object of consciousness in infancy. In: Rochat, P. (ed) The self in infancy: Theory and research. Elsevier: Amsterdam, pp. 35–52.

Caulfield, H. J. (2002) Review of the conscious mind: In search of a fundamental theory. Human Nature Review 2:242–243.

Caulfield, H. J./Johnson, J. L./Schamschula, M. P./Inguva, R. (2001) A general model of primitive consciousness. Cognitive Systems Research 2:263–272.

Chalmers, D. (1996) The conscious mind: In search of a fundamental theory. Oxford University Press: New York.

Churchland, P. S. (1989) Neurophilosophy. MIT Press: Cambridge MA.

Dennett, D. (1991) Consciousness explained. Little, Brown: Boston MA.

Eccles, J. (1991) The evolution of consciousness. Taylor & Francis: London.

Fichte, J. G. (1794–1795) Über den Begriff der Wissenschaftslehre [On the concept of the science of knowledge].

Freeman, W./Watts, J. W. (1941) The frontal lobes and consciousness of self. Psychosomatic Medicine 3:111–119.

Gallagher, S./Shear, J. (eds) (2000) Models of the self. Imprint Academics: New York.

Hume, D. (1739) A treatise of human nature. London.

James, W. (1890) The principles of psychology. Holt: New York.

LeDoux, J. (2002) Synaptic self: How our brains become who we are. Viking: New York.

MacPhail, E. (1998) The evolution of consciousness. Oxford University Press: New York.

Mitchell, R. W. (1993) Mental models of mirror self-recognition: Two theories. New Ideas in Psychology 11:295–325.

Penrose, R. (1987) Quantum physics and conscious thought. In: Hiley, B./Peat, D. (eds) Quantum implications: Essays in honour of David Bohm. Routledge & Kegan Paul: New York, pp. 105–120.

Penrose, R. (1999) The emperor's new mind. Oxford University Press: New York. Originally published in 1987.

Power, N. P. (2001) The origins of self-consciousness. Minds and Machines 11:133–137.

Stapp, H. P. (1998) The evolution of consciousness. In: Hameroff, S./Kaszniak, A./Scott, A. (eds) Toward a science of consciousness II. MIT Press: Cambridge MA.

Strawson, G. (1997) The self. Journal of Consciousness Studies 4:405–428.

James H. Fetzer

Evolving Consciousness: The Very Idea!

Abstract

Discovering an adequate explanation for the evolution of consciousness has been described as "the hard problem" about consciousness that we would like to understand. This difficulty becomes compounded by the introduction of such notions as the unconscious or the preconscious as its counterparts, at least for species of the complexity of human beings. An evaluation of the prospects for unconscious factors as exerting causal influence upon human behavior, however, depends upon understanding both the nature of evolution and the nature of consciousness. This paper sketches a theoretical framework for understanding both phenomena in general with regard to their various forms and suggests the evolutionary function of consciousness in genetic and in cultural contexts. It becomes increasingly apparent that, given a suitable conceptual framework, the evolution of consciousness might not be such a "hard problem", after all.

Key words

Consciousness, cognition, evolution, stimuli, signs, minds, semiotic systems, awareness, self-awareness, communication, signals.

The "Hard" Question

Philosophers spend most of their time dealing with vague and imprecise notions, attempting to make them less vague and more precise (FETZER 1984). When we are dealing with notions like "the unconscious mind", where we have only a vague notion of consciousness and an imprecise notion of the mind, it may be appropriate to propose a few suggestions in an effort to sort things out a bit better, especially when the role of evolution in producing mentality and consciousness appears to be poorly understood. This study attempts to shed light on these problems by exploring how consciousness of different kinds might contribute to evolution in relation to its causal mechanisms.

Why did consciousness evolve? has been called "the hard problem" and some have even denied that it can be an adaptation (HARNAD 2002). So*What are the adaptive benefits of consciousness?* and *How does consciousness enhance the prospects for survival and reproduction of species that possess it*? are crucial questions. But the correct answers necessarily depend upon the nature of consciousness itself. In his *Kinds of Minds*, for example, Daniel DENNETT (1996) suggests that consciousness is sensitivity plus some additional factor "x", yet he thinks there might be no such "x". If it is merely the capacity for sensation and sensation is no more than a propensity for undergoing change, then consciousness might even be separate from mentality. There might be nothing distinctive about consciousness nor motive for its evolution.

If consciousness is instead a sensory awareness of the sensible qualities of things, such as their colors, shapes, sizes, by comparison, it might make a difference and even imply the presence of minds. In *The Evolution of Culture in Animals*, for example, John BONNER (1980) describes *E. coli* bacteria as moving toward 12 chemotactic substances and as moving away from 8 others. Assuming the ones it moves toward are nutrient or beneficial, while the ones it moves away from are harmful or deleterious, it is not very difficult to imagine how evolution could have produced this result at this stage for those bacteria. Perhaps "the hard problem" might turn out not to be a hard problem, after all.

The "Black Box" Model

We tend to operate on the basis of a rather simple model—a "black box" model—for organisms. We have a stimulus S that brings about a response R by an organism O with a certain probability or propensity p (FETZER 1981, 1993a). The propensity p for response R by an organism O, when subject to stimulus S, can be formalized as, Stimulus S ==> [Organism O =p=> Response R]

Figure 1. The black box.

or, alternatively, by exchanging the positions of the organism O and the stimulus S,

Organism O ==> [Stimulus S =p=> Response R]

Figure 2. The black box (reversed).

where different species and different organisms of the same species may be subject to different ranges of stimuli S and ranges of response R with different propensities.

This model does not offer any analysis of processes internal to O, which makes it a "black box" model. A more refined analysis, however, takes into account the possible existence of links that relate an initial INTERNAL response R1 to the occurrence of one or more possible additional INTERNAL responses Ri, where these responses may lead to EXTERNAL responses Rj of motion or sound by the organism formalized as follows:

(EXTERNAL)	O ==>
(INTERNAL)	[(S =p1=> R1) & (R1 =p2=> R2) &
	(R2 =p3=> R3) & …] =pj=>
(EXTERNAL)	Rj

Figure 3. A more refined model.

Thus, for an ordinary organism of kind O, under suitable circumstances, an external stimulus S, which might be a sight or a sound, causes a pattern of neural activation R1, which in turn may (probabilistically) bring about a pattern of neural activation R2, . . . , which in turn may (probabilistically) bring about other patterns of neural activation, which may eventually lead to (public) external responses Rj, such as motion or sounds. The simpler the organism, the simpler these internal links (Fetzer 1990, 1996, forthcoming).

This approach invites the introduction of at least three measures of complexity that could distinguish between species or even conspecifics as members of the same species, based upon various properties of such links as possible internal causal chains, namely: (a) the complexity of these internal causal chains, especially with regard to (i) number of possible links and (ii) their deterministic or probabilistic character; (b) the temporal interval between the initial stimulus S and the ultimate behavioral response R, if any; and (c) the complexity of those possible responses that organisms display themselves.

Human Behavior

A simple example in the case of human behavior might be making a date, such as to attend this conference. We may do so months in advance, but our behavioral responses to our commitments are only displayed when the time draws near. This reflects the consideration that human behavior arises as a result of the complex causal interaction between multiple factors of the kinds *motives, beliefs, ethics, abilities,* and *capabilities*, where behavior may be a probabilistic manifestation of their interaction:

MOTIVES (m1, m2, … mn) &		
BELIEFS (b1, b2, … bn) &		RESPONSES r1, r2, …rn,
ETHICS (e1, e2, … en) &	=p=>	including motion and
ABILITIES (a1, a2, … an) &		sounds
CAPABILITIES (c1, c2, … cn)		

Figure 4. Human behavior as a probabilistic effect.

The success or failure of the actions we undertake on the basis of those motives and beliefs, however, depends very heavily on our *opportunities* as the way things are, which reflects the truth and the completeness of our beliefs (Fetzer 1989, 1990).

While one mental state may bring about another mental state through a series of transitions between links of the kind described above, the totality of factors that interact to (probabilistically) bring about our behavior consists of specific values of variables of each of these kinds, where one complete set of values for the variables *motives, beliefs, ethics, abilities,* and *capabilities* constitutes a *context*. The concept of a context turns out to be fundamental to meaning and mind (Fetzer 1991, 1996).

The difference between deterministic and indeterministic behavior can then be spelled out as follows. Relative to a context, when the same behavior would occur in every case without exception, then that behavior is *deterministic*. When one or another behavior within a fixed class would occur in every case without exception, with a constant probability, then that behavior is *indeterministic*. Consequently, even persons in the same context C can manifest different behavior so long as it is among the possible outcomes that occur with a fixed propensity within that context.

With regard to motives, for example, if you like Heavenly Hash twice as much as you do Peppermint BonBon, where they are your clear preferences in ice cream, then we would expect that you would choose Heavenly Hash about twice as often as Peppermint BonBon when you enter Baskin Robbins.

You would not know which you would pick on any single visit, but over time you would pick one about twice as often as you pick the other. Frequencies are produced by propensities across trials, which can explain them and for which they function as evidence (FETZER 1981, 1993a, 2002a).

Meaning and Behavior

What holds for motives also holds for beliefs, ethics, and the other variables that affect our behavior. With regard to beliefs, for example, I happen to live at 2021 E. 4th Street in Duluth, MN. If someone were to believe instead that I lived at 2017, that would have multiple manifestations in their behavior, such as the directions they might give to get to my house, what they would write on a letter they wanted to mail to me, where UPS and FED/EX deliveries to me would be made, and the like.

This approach supports a dispositional theory of meaning, according to which *the meaning of a belief, Bi*, is the difference that Bi makes over alternatives Bj relative to every context consisting of specific values of motives, of other beliefs, and so forth, where, when there is no difference in the totality of behavior that would be displayed given Bj as opposed to Bi across every context, then the meaning of Bj is the same as Bi (FETZER 1991, 1996). And it turns out that meaning itself is amenable to degrees.

Those who know that my home is the fourth house on the block on the high side, for example, might be able to find it without great effort because of their other beliefs about how to get around in Duluth, but for other purposes the street number would be required. Some but not all of the same behavior would result from those overlapping beliefs. Two half-dollars, four quarters, and so on has the same purchasing power as a dollar, but in some contexts carrying a bill rather than bulky change might matter.

This account of meaning, which connects stimuli S with responses R by means of internal dispositions of an organism O, comports with a theory of concepts and even of mind. If we think of *concepts* as constellations of habits of thought and habits of action, then when an experience is subsumed by means of a concept, the expectable outcome is whatever behavioral effects would (probably) be produced in a context. Some concepts will be innate, while others may be acquired (FETZER 1991, 1996).

Another species that exemplifies these notions is that of vervet monkeys, which makes at least three different kinds of alarm calls. In his *Introduction to Ethology,* P. J. B. SLATER (1985) reports that one such call warns of a land-borne predator in the vicinity and, when the monkeys hear this call, they climb up into the trees to evade it. Another warns for an air-borne predator in the vicinity and, when they hear it, they crawl down under the bushes for protection. And the third is for things on the ground, where they climb down and poke around so they can see just what is going on.

Our behavior, especially voluntary, turns out to be a partial manifestation of meaning to us, where the meaning of meaning to us turns out to be the multiple potentialities for behavior in the presence of something S, where I want to identify that S more precisely as *a stimulus of a certain special kind*, which makes a crucial difference to our behavior. The suggestion I am going to make is that an approach, which has not received a lot of attention as yet, but that was advanced by Charles S. PEIRCE—whom I consider to be the only great American philosopher—can help to clarify and illuminate the nature of mind.

The Nature of Signs I

According to PEIRCE, *a sign* is a something that stands for something else in some respect or other for somebody. A simple illustration is a red light at an intersection. For qualified drivers who know the rules of the road, that light stands for applying the breaks and coming to a complete halt, only proceeding when the light changes and it is safe to do so. Under ordinary circumstances—in a "standard context", let us say—that is precisely the behavioral manifestation that we expect to occur (FETZER 1988, 1991).

This would be an example of an appropriate behavioral response for somebody who understood the rules of the road and is not incapacitated from exercising that ability, as might be the case if, for example, they were blindfolded. And of course there can be other signs with the same meaning, such as, in this case, a stop sign or an officer with his palm extended, which have essentially the same meaning (of applying the breaks and coming to a complete halt, but only proceeding when the officer tells you to do so). PEIRCE called the complex of dispositions of a user to respond to a sign its "interpretant".

PEIRCE suggests there are three different ways in which signs can be "grounded" or related to those things for which they stand. The first is on the basis of *resemblance relations*, where the sign looks like (tastes like, smells like, feels like, or sounds like) that for which it stands. Examples include statues, photographs, and paintings, when realistically construed. (Thus, PICASSO achieved a niche in the history of art when he violated the canons of representation of the nude female.) PEIRCE called these "icons".

My driver's license exemplifies an important point about icons. As you might or might not be able to see, my license photo looks a lot like me—maybe on not such a great day—but if you turn it on its side, it no longer resembles me, because I am just not that thin. What this implies is that even the use of the most basic kind of sign, an icon, presupposes *a point of view.* Anyhing incapable of a point of view, therefore, is incapable of using signs or of possessing a mind, a point to which we shall return.

The second mode of grounding that PEIRCE introduced is *causal relations*, where a cause stands for its effects, effects stand for their causes, and so forth. Thus, smoke stands for fire, fire for smoke, ashes for fire, and so on, while red spots and elevated temperature stand for the measles—which means that there may be special classes of individuals who are practiced in reading signs of certain kinds, such as scientists and physicians, but also those whose parallel claims may be suspect, such as palm readers and crystal-ball gazers. PEIRCE called these signs "indices" (as the plural for "index").

The Nature of Signs II

The third mode of grounding PEIRCE introduced involves mere *habitual associations* between signs and that for which they stand, where the most familiar examples are the words that occur in ordinary languages, such as "chair" and "horse" in ordinary English. These words certainly do not look like or resemble nor are they causes or effects of that for which they stand. Unlike icons and indices, which might be thought of as "natural" signs because they are there in nature whether we notice them or not, these signs are ones we have to make up or to create. These "artificial" signs are known as "symbols".

In order for a specific something to stand for something else in some respect or other for somebody on a specific occasion, that somebody must have the ability to use signs of that kind, s/he must not be incapacitated from exercising that ability, and that sign must stand in an appropriate causal relationship to that sign user. If a red light were invisible to a driver because of a driving rain (a dense fog, overgrown shrubbery, or whatever), it could not exert its influence on that sign user on that occasion any more than if s/he had been temporarily blinded by a flash of lightning or an oncoming car (FETZER 1990, 1996).

Even more interesting, perhaps, is the realization that the specific something for which something stands in some respect or other need not exist. We can have signs for persons who do not exist, such as Mary Poppins and Santa Claus, or for species of things, such as unicorns and vampires, that do not exist, without incapacitating those signs from standing for things of those kinds. We can even make movies about alien visitations and American werewolves in London. Which means that the use of signs has enormous scope and range with respect to those things for which they can stand. They do not even have to exist!

The Nature of Minds

The sign relationship, therefore, is three-placed (or "triadic"), where a something, S, stands for something else, x, (in some respect or other) for somebody, z. The meaning of a sign is then the totality of causal influences it would exert across possible contexts, Ci:

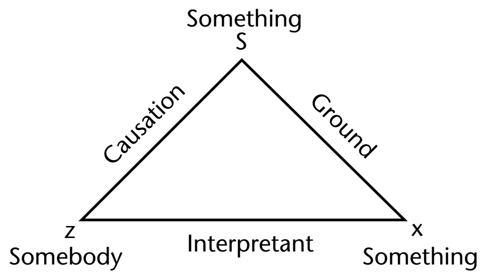

Figure 5. The triadic sign relationship.

When we pause to consider more precisely the kind of thing for which something can stand for something else, however, it becomes extremely attractive to entertain the hypothesis that *the capacity to use signs* might be exactly what distinguishes minds.

Let us focus on the sign user z rather than the sign S and avoid taking for granted that the kinds of things for which something can stand for something else have to be human by abandoning the term "somebody" and use the more neutral term "something". Then anything, no matter whether it happens to be human being, (other) animal, or inanimate machine, for which something (a sign) can stand for something else in some respect or other *possesses a mind.* And let us refer to systems of this kind as things that are capable of using signs as *semiotic systems* (FETZER 1988, 1989, 1990).

Semiotic Systems

"Interpretant" thus stands for a system's semiotic dispositions as the totality of ways it might respond (probabilistically) to the presence of a sign within different contexts. Its behavior in context Ci can therefore differ from its behavior in Cj in the presence of the same sign (FETZER 1991). And a semiotic system z can be diagrammed as follows:

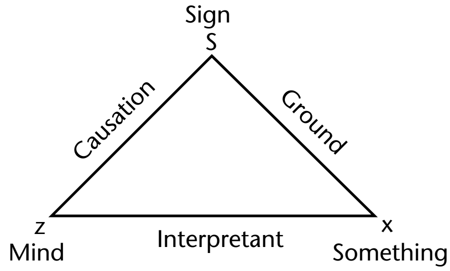

Figure 6. A semiotic system.

The grounding relations between signs and that for which they stand (by virtue of relations of resemblance, of cause-and-effect, or of habitual association, as we have discovered), are therefore crucial to the nature of semiotic systems. Unless that causal connection between the presence of something and the (potential or actual) behavior of a system obtains *because* it functions as an icon, an index, or a symbol for that system (by virtue of its grounding relation of resemblance or of causation or of habitual association), it cannot be a *semiotic* connection (FETZER 1990, p278).

Semiotic systems for which things function as signs afford a basis for separating systems that have minds from others that do not, such as digital machines, which lack the grounding relationship relating signs to those things for which they stand. This difference can also be diagrammed to display this crucial difference as follows:

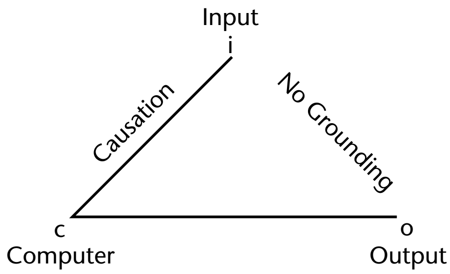

Figure 7. An input–output system.

Thus, although they are designed to process marks on the basis of their shapes, sizes, and relative locations, those marks mean nothing to those digital machines, say, as inventories or as dollars and cents. They should therefore be characterized, not as semiotic systems, but as input/output systems instead, where the inputs that exert causal influence upon them are properly understood to function merely as stimuli rather than as signs. They can be called "symbol systems", provided that does not imply that they use symbols in PEIRCE's sense (FETZER 1988, 1990, 1996, 2002b).

Communication and Convention

Another important distinction that can be drawn is that communication between semiotic systems is promoted when those systems use signs in similar ways. When a sign-using community reinforces common sign-using behavior by means of some system of institutions, such as schools, those customs, traditions, or practices take on the status of *conventions*, which promotes the objectives of communication and cooperation, thereby facilitating the pursuit of community goals (FETZER 1989, 1991).

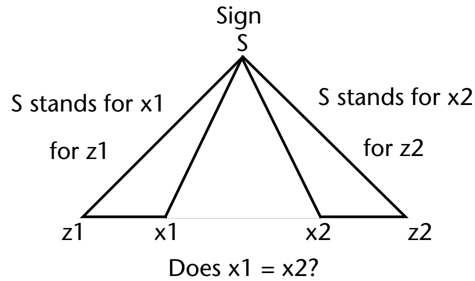

Figure 8. Communication situations.

When one semiotic systems uses signs to communicate with another semiotic system, then those signs assume the character of *signals*. There thus appears to be a hierarchy between mere stimuli, signs, and signals, because every signal is a sign and every sign is a stimulus, but not vice versa. Causes that can produce changes in inanimate objects, for example, are stimuli but not signs, just as things that stand for other things are signs for those systems even if they are not signals. While all three—stimuli, signs, and signals—are possible causes that can affect the behavior of different systems, only signs and signals entail the presence of minds.

Consciousness and Cognition

Even more important, however, the theory of minds as semiotic systems also provides illuminating conceptions of consciousness and of cognition, where both turn out to be adequately defined only relative to signs of specific kinds. Thus, a system z is *conscious (with respect to signs of a specific kind S)* when (a) z has the ability to use signs of kind S and (b) z is not incapacitated from using signs of that kind within its present context C. *Cognition (with respect to a specific sign of kind S)* thus occurs as the effect of a causal interaction between a system z and a sign S when (a) z is conscious with regard to signs of kind S and (b) a sign of kind S occurs in suitable causal proximity to z, which brings about the activation of its mental states as the outcome of a suitable opportunity (FETZER 1989, 1990, 1996).

Consciousness (with respect to signs of kind S) = df ability + capability (within a context)

Cognition (of a specific sign of kind S) = df an effect of consciousness + opportunity

Figure 9. Consciousness and cognition (informal).

The conception of minds as semiotic systems (sign-users) thus not only brings with it the definition of mentality as semiotic ability but useful conceptions of consciousness and of cognition. Informally expressed, consciousness (with respect to signs of specific kinds) combines the *ability* to use signs of that kind with the *capability* to exercise the ability, while cognition (relative to a specific sign) combines *consciousness* with respect to signs of that kind and the *opportunity* for causal interaction with a sign of that kind. That definition can be complemented with a general criterion of mentality, which is the capacity to make a mistake, since anything that can make a mis-take has the ability to take something to stand for something, which is the right result (FETZER 1988, 1990).

The outcome of this approach is the introduction of a theory of mentality that is applicable to human beings, to (other) animals, and to inanimate machines, if such a thing is possible. It yields a system of types of minds of increasing strength, from iconic to indexical to symbolic, where symbolic presupposes indexical and indexical iconic, but not vice versa. These types and criteria of their presence are shown in Figure 10, where an evidential indicator of the presence of iconic mentality is the capacity for *type/token recognition* of instances as instances of specific kinds; of indexical is *classical PAVLOVIAN conditioning* as the generalization of a cause inducing an effect; and of symbolic mentality *SKINNERIAN operant conditioning*, where one thing comes to stand for another based merely upon habitual association (FETZER 1988, 1990).

	Mentality		
	Type I	**Type II**	**Type III**
Definition	iconic	indexical	symbolic
Criterion	type/token recognition	classical PAVLOVIAN conditioning	SKINNERIAN operant conditioning

Figure 10. Basic modes of mentality.

Higher Modes of Mentality

This approach invites the evolutionary hypothesis that various biological species are predisposed toward mentality of specific types, which would be expected to be distributed as a reflection of their evolutionary standing, the lowest organisms with the lowest levels of mentality, the higher with higher. Indeed, there appear to be at least two higher modes of mentality that are characteristic of human beings, which are the capacity to fashion arguments as *transformational mentality* and the ability to use signs to stand for other signs as *metamentality*, especially for the purpose of criticism, where sign-users can subject signs to changes intended to improve them, as Figure 11 displays.

Among the virtues of the conception of minds as semiotic systems is that it allows for the existence of modes of mentality that are less sophisticated than those involved in the use of language, which appears to be a relatively late phenomenon in evolution (DONALD 1991; FETZER 1993b, 1993c). The extraordinary attention to which it has been subject by Noam CHOMSKY's work on grammar as a species-specific innate syntax and Jerry FODOR's work on meaning as a species-specific innate semantics has reached its latest incarnation in work such as that of Stephen PINKER (1997), who holds that the human mind is a computer for survival and reproduction, and of Euan MACPHAIL (1998), who maintains the key to the evolution of consciousness is the evolution of language.

	Higher Mentality	
	Type IV	**Type V**
Definition	transformational	metamentality
Criterion	logical reasoning	criticism

Figure 11. Higher modes of mentality.

If the evolution of language were the key to the evolution of consciousness, then, insofar as language is a phenomenon relatively late in evolution, it would be rather difficult to imagine how consciousness could have evolved at all. Preoccupation with language truncates consideration of multiple modes of meaning and non-human kinds of minds. Not only are iconic and indexical mentality more primitive than symbolic, but preoccupation with linguistic transformations and syntactical structures manages to focus on higher modes of mentality to the neglect of lower, while even placing the syntactic cart before the semantic horse. As Thomas SCHOENEMANN has argued (and I agree), that syntax evolved as an emergent response to semantic complexity affords a better explanation of the phenomena than its innate alternatives (SCHOENEMANN 1999).

Conceptions of Consciousness

The idea that the mind is a computer that evolved through natural selection, of course, takes for granted that, at some appropriate level of description, both minds and machines operate on the basis of the same or similar principles, which already appears to be false given the difference in grounding relations. But modeling minds after machines also confounds languages as products of the evolution of culture with species as products of the evolution of genes. The relative adequacy of alternative theories (of consciousness, mentality, and language) may be assessed by extent to which they are able to explain the full range of related phenomena (of consciousness, mentality, and language), where, I would submit, the semiotic conception encounters no serious rivals.

As an illustration, consider the multiple modes of consciousness that can be differentiated within the scope of this approach. Those that do not make reliance upon signs lack the semiotic dimension distinctive of mentality. The DENNETT hypothesis that consciousness may be nothing more than sentience qualifies thermostats, litmus paper, and thermometers as "conscious", yet is not sufficient to endow them with mentality (FETZER 1997). They are thus examples of sensitivity as the susceptibility to stimuli that does not implying mentality as a species of "consciousness" without minds. Let's call this (C-1).

A stronger mode of consciousness would combine sensitivity with semiotic ability, which implies the presence of mind. Call this (C-2). A third mode of consciousness would combine semiotic ability with self-awareness involving the use of signs to stand for the sign-user itself. Call this (C-3). Yet a fourth mode of consciousness would combine self-awareness with the capacity for articulation, which we shall call (C-4). A fifth mode of consciousness would combine self-awareness with the capacity for articulation and the ability to communicate with others using signs as signals. Let us call this final mode (C-5).

(C-1)	**Sensitivity**
	stimuli with causal influence but does not imply mentality: thermostats, thermometers, litmus paper as a kind of mindless consciousness
(C-2)	**Semiotic ability**
	sensitivity regarding stimuli that stand for something in some respect for something; hence, (C-2) implies (C-1) and the presence of mind
(C-3)	**Self-awareness**
	semiotic ability that includes signs that stand for the sign user itself for the sign user; so (C-3) implies (C-2) with self-referential ability
(C-4)	**Self-awareness with articulation**
	semiotic ability that includes signs that stand for the user itself with the ability to articulate that self-awareness; so (C-4) implies (C-3) with articulative ability
(C-5)	**Self-awareness with capacity for communication**
	semiotic ability that includes signs standing for oneself and other conspecifics, which promotes cooperation, so (C-5) implies (C-4) with signals

Figure 12. Five modes of consciousness.

This schema does not represent the only possible kinds of consciousness but rather serves as a template to consider the prospective roles of consciousness in evolution. In this case, for example, each mode of consciousness implies each of the lower modes, where (C-5) implies (C-4), (C-4) implies (C-3), and so forth. If there are cases of communication involving signals, which presumably would be at the level of (C-5), such as vervet monkey alarm calls, where their use of signals may or may not be accompanied by self-referential ability at the level of (C-3), then it has exceptions that would display the desirability of deviant typological schemes.

Evolution and Consciousness

Evolution understood as a biological process should be characterized in terms of three principles, namely: that more members are born live into each species than survive to reproduce; that crucial properties of offspring are inherited from their parents; and that several forms of competition between the members of a species contribute to determining which of them succeeds in reproducing. The mechanisms that tend to produce genetic variation include genetic mutation, sexual reproduction, genetic drift, and genetic engineering, while the mechanisms that tend to determine which members of existing populations tend to survive and reproduce are natural selection, sexual selection, artificial selection, and group selection (FETZER 2002c).

The question with which we began, you may recall, was, *Why did consciousness evolve?*, which is amenable to alternative formulations that include, *What are the adaptive benefits of consciousness?* but also *How does consciousness enhance the prospects for survival and reproduction of species that possess it*? Having clarified the nature of consciousness sufficiently to make these questions meaningful (or at least interesting) enough to pursue them, the objective becomes to consider each of these causal mechanisms in turn to ascertain whether consciousness of any of these five modes would provide adaptive benefits in order to answer "the hard question".

The following table reflects the big picture, in general, as the intersection of the eight different evolutionary mechanisms with those modes of consciousness that might enhance them or benefit from them. The first four are modes that promote variability in the gene pool. Consciousness beyond sensitivity would appear to make no difference to the occurrence of genetic mutation, which of course presupposes consciousness (C-1). Similarly for sexual reproduction and genetic drift, understood as causal processes apart from the mechanisms that determine who mates with whom and under what conditions.

Mechanism	Consciousness
(1) Genetic mutation	(C-1)
(2) Sexual reproduction	(C-1)
(3) Genetic drift	(C-1)
(4) Genetic engineering	(C-5)
(5) Natural selection	(C-1) to (C-5)
(6) Sexual selection	(C-2) to (C-5)
(7) Group selection	(C-5)
(8) Artificial selection	(C-5)

Figure 13. Adaptive roles of modes of consciousness.

Genetic engineering, by contrast, requires highly sophisticated mental abilities that would appear to benefit from reasoning skills and critical thinking up to the level of (C-5). The emergence of consciousness at levels far beyond (C-1) would provide adaptive benefits. In the case of natural selection, all these modes would be beneficial in competition with conspecifics for food and other resources. Success in sexual selection, moreover, would benefit from self-referential abilities and the capacity for articulation, not to mention the ability to transmit appropriate signals. Artificial selection and group selection could not operate without communication.

If these considerations are well-founded, then they suggest that the potential adaptive benefits of consciousness are both obvious and profound. In response to the question, different modes of consciousness appear to enhance the prospects for survival and reproduction by species that possess them. Intriguingly, the motives for consciousness to evolve differ in relation to different evolutionary mechanisms. It should come as no surprise that natural selection and sexual selection should both benefit from consciousness up to the highest kinds, where genetic engineering and artificial and group selection could not function without consciousness around (C-5).

Minds Are not Machines

What this exercise has secured is a plausibility proof that evolution can produce consciousness among its varied manifestations, since organisms with these kinds of abilities would secure advantages in competition with nature and with conspecifics across a wide range of evolutionary mechanisms. This means that there would be adaptive benefits from possessing consciousness of these various kinds that would enhance the prospects for survival and reproduction among those possessing them. It should also be observed, however, that this analysis could be improved upon by, for example, systematically integrating consideration for different kinds of minds.

There should not be much room for doubt, for example, that higher modes of consciousness tend to presume higher types of mentality, where transformational mentality and metamentality can greatly extend the abilities of organisms in dealing with conspecifics and their environments. All of this may even seem to reinforce the claim that the human mind is a computer for survival and reproduction. That claim, however, trades upon an ambiguity. There is some general sense in which the human mind is a processor for survival and reproduction, but this, from the point of view of evolution, is a trivial claim. The sense in which the human mind is a computer, alas, implies that they operate on the basis of the same or similar principles, which is false.

We have already seen that digital machines lack a grounding relation that typifies not just human minds but every mind. (Compare Figure 7 with Figure 6.) So that is one important difference, which we might call "the static difference". Another is that these machines function on the basis of *algorithms* implemented by using programs, which execute operations in specific sequences of steps. They have definite starting points and definite stopping points, where their application is perfectly general and they always yield a correct solution to a problem in a finite number of steps. When you think about it, these are important differences between computing and thinking.

How many kinds of thinking have these properties? Certainly neither perception nor memory nor dreams or daydreams come close. None of them ordinarily qualifies as solving problems. None of them has a definite starting point and another definite stopping point. None of them can be counted upon to yield correct solutions in finite steps. We might call this "the dynamic difference". What this means is that digital machines and human beings do not function on the basis of the same or even similar principles. They exemplify the static difference and the dynamic difference, which means that they are systems of distinctly different kinds. Human beings surely are systems for survival and reproduction, but that does not turn them into computers. PINKER is wrong, because minds are not machines (FETZER 1990, 1994, 1996, 2002b).

Genetic vs. Cultural Evolution

In an earlier book, Steven PINKER (1994) embraced the hypothesis of a uniquely human "language instinct", while acknowledging that this species-specific conception does not appear to be logically compatible with modern Darwinian theory of evolution "in which complex biological systems arise by the gradual accumulation over generations of random genetic mutations that enhance reproductive success" (p333). His solution is to explain that the history of evolution produces a bushy structure, not an ordered sequence, where his account is not endangered by its incapacity, for example, to show that monkeys have language. But surely it would be more reasonable to suppose that our evolutionary relatives, including monkeys, have some counterpart ability to use different yet comparable methods for communication. A broader semiotic framework would relate the use of signs to the subsumption of experience by means of concepts.

An adequate understanding of the evolution of language and mentality, moreover, heavily depends upon a firm grasp of the differences between genetic and cultural evolution. By adopting the common distinction between "genes" as units of genetic evolution and "memes" as units of cultural evolution, John BONNER (1980) already identified three important differences, because (1) genes can exist independently of memes, but not conversely (there are no disembodied thoughts); (2) genes are transmitted but once per organism, while memes can be acquired over and over; and, (3) that the rate of change for genes is constrained by gestation, whereas the rate of change for memes approximates the speed of information transmission. Thus,

Genetic Evolution	Cultural Evolution
(1) Genes can exist independently of memes	(1') Memes cannot exist independently of genes
(2) One time transmission of information (conception)	(2') Multiple opportunities for information transmission
(3) Changes very slow (bound by rate of reproduction)	(3') Changes very fast (bound by speed of light)

Figure 14. Genetic vs. cultural evolution (BONNER).

Other differences distinguish them as well, however, which in some contexts may be even more important. Thus, for example, the genetically heritable properties of organisms are ones that any organism with those genes could not be without (given fixed environmental factors) as permanent properties, while the memetic properties of organisms are often transient and acquired. The causal mechanisms underlying cultural evolution are rooted in the semiotic abilities of the species (FETZER, forthcoming).

Ultimately, distinctions must be drawn between species for which their mental abilities are innate, inborn, and species-specific, and those for which their mental abilities can be enhanced through conditioning, learning, and even critical thinking. Low-level species, such as bacteria, may satisfy the conception of evolution where complex biological systems arise by the gradual accumulation over generations of random genetic mutations that enhance reproductive success. But other species far transcend the limitations that those constraints would impose. The only permanent properties related to language that humans have to possess are predispositions for the acquisition of concepts as habits of thought and habits of action, including the use of icons, indices, and symbols. There is no need for a "language instinct" as an innate disposition to use language (FETZER 1991, SCHOENEMANN 1999, DUPRÉ 1999).

(4) affect permanent properties	(4') affect merely transient properties
(5) mechanisms of genetic change are Darwinian, including: genetic mutation natural selection sexual reproduction … artificial selection genetic engineering	(5') mechanisms of memetic change are Lamarckian, including: classical conditioning operant conditioning imitating others … logical reasoning rational criticism

Figure 15. Genetic vs. cultural evolution (FETZER).

Concluding Reflections

In a broader sense, thinkers like PINKER, FODOR, CHOMSKY, and MACPHAIL, who are preoccupied with language, have missed the boat by taking syntax to be more basic than semantics. When it comes to evolution, they have some general appreciation for the origin of species but little understanding of key differences between genetic and cultural evolution. They have developed their theories largely independently of the question, *But where did language come from?*, as though it could arrive on the scene full-blown as a language of thought rich enough to sustain every sentence in every language—past, present, or future—that did not have to be a product of evolution!

The considerations adduced here, however, provide a fertile point of departure for other studies that carry this approach into new domains. While the theory of minds as semiotic systems clarifies and illuminates the very idea of consciousness as an evolutionary phenomenon, the elaboration of that approach for unconscious and preconscious phenomena requires further exploration. At the very least, it makes clear that mental phenomena are semiotic phenomena involving the use of signs. When organisms are exposed to stimuli for which they lack corresponding concepts, for example, then they are unable to subsume them and remain merely "preconscious". When they subsumed by concepts for which those organisms have no signals, they are restricted to private use and might be said to be "unconscious".

This raises the possibility that the notions of "preconscious" and of "unconscious" may ultimately be envisioned as *relative* to kinds of consciousness. The study of Freud should contribute considerably within this context, since no one ever had a firmer grasp of the intricacies of the human mind with regard to its conscious, unconscious, and preconscious dimensions (SMITH 1999). Although the semiotic conception elaborated here supports appealing accounts of consciousness and of cognition, which have obvious evolutionary implications for the origin of species, its implications for the preconscious and unconscious invite future development.

The theory of minds as semiotic systems presents an attractive alternative to models of the mind inspired by computers and language. Their respective merits should be assessed on the basis of the criteria of comparative adequacy for scientific theories, which include (a) the clarity and precision of the language in which they are couched; (b) their respective scopes of application for explaining and predicting the phenomena to which they apply; (c) their respective degrees of confirmation on the basis of suitable observations, measurements, and experiments; and (d) the simplicity, economy, or elegance with which their scopes of application happen to be attained (FETZER 1981, 1993a). By this standard, the semiotic approach, which applies to humans, (other) animals, and even machines, if such a thing is possible, provides a far superior framework for understanding consciousness and cognition, including its ability to place "the hard problem" in proper evolutionary perspective.

Author's address

James H. Fetzer, Department of Philosophy, University of Minnesota, Duluth, MN 55812, USA. Email: jfetzer@d.umn.edu

References

Bonner, J. (1980) The evolution of culture in animals. Princeton University Press: Princeton NJ.

Dennett, D. (1996) Kinds of minds. Basic Books: New York NY.

Donald, M. (1991) Origins of the modern mind. Harvard University Press: Cambridge MA.

Dupré, J. (1999) Pinker's how the mind works. Philosophy of Science 66:489–493.

Fetzer, J. H. (1981) Scientific knowledge. Reidel: Dordrecht.

Fetzer, J. H. (1984) Philosophical reasoning. In: Fetzer, J. (ed) Principles of philosophical reasoning. Rowman & Allanheld: Totowa NJ, pp. 3–21.

Fetzer, J. H. (1988) Signs and minds: An introduction to the theory of semiotic systems. In: Fetzer, J. (ed.), Aspects of artificial intelligence. Kluwer: Dordrecht, pp. 133–161.

Fetzer, J. H. (1989) Language and mentality: Computational, representational, and dispositional conceptions. Behaviorism 17(1):21–39.

Fetzer, J. H. (1990) Artificial intelligence: Its scope and limits. Kluwer: Dordrecht.

Fetzer, J. H. (1991) Primitive concepts. In: Fetzer, J. et al. (eds.) Definitions and Definability. Kluwer: Dordrecht.

Fetzer, J. H. (1993) Philosophy of science. Paragon: New York NY.

Fetzer, J. H. (1993a) Donald's Origins of the modern mind. Philosophical Psychology 6(3):339–341.

Fetzer, J. H. (1993b) Evolution needs a modern theory of the mind. Behavioral and Brain Sciences 16(4):759–760.

Fetzer, J. H. (1994) Mental algorithms: Are minds computational systems? Pragmatics and Cognition 2(1):1–29.

Fetzer, J. H. (1996) Philosophy and cognitive science (2nd ed). St. Paul: Paragon MN.

Fetzer, J. H. (1997) Dennett's kinds of minds. Philosophical Psychology 10(1): 113–115.

Fetzer, J. H. (2002a) Propensities and frequencies: Inference to the best explanation. Synthese 132(1–2), pp. 27–61.

Fetzer, J. H. (2002b) Computers and cognition: Why minds are not machines. Kluwer: Dordrecht.

Fetzer, J. H. (2002c) Introduction. In: Fetzer, J. (ed) Consciousness evolving. John Benjamins: Amsterdam, pp. xiii–xix.

Fetzer, J. H. (forthcoming) The Evolution of intelligence: Are humans the only animals with minds? Chicago IL: Open Court.

Harnad, S. (2002) Turing indistinguishability and the blind watchmaker. In: Fetzer, J. (ed) Consciousness evolving. John Benjamins: Amsterdam, pp. 3–18.

MacPhail, E. M. (1998) The evolution of consciousness. Oxford: New York NY.

Pinker, S. (1994) The language instinct. William Morrow: New York NY.

Pinker, S. (1997) How the mind works. W. W. Norton: New York NY.

Schoenemann, P. T. (1999) Syntax as an emergent characteristic of the evolution of semantic complexity. Minds and Machines 9(3):309–334.

Slater, P. J. B. (1985) An introduction to ethology. Cambridge University Press: Cambridge UK.

Smith, D. L. (1999) Freud's philosophy of the unconscious. Kluwer: Dordrecht.

Temple Grandin

Do Animals and People with Autism Have True Consciousness?

Abstract

Some philosophers think animals are not conscious because they do not have language. I am autistic and I think in pictures. If the philosophers are correct, I would have to conclude that I am not conscious.
Language is used to narrate the visual images that form my thoughts. I think consciously with the part of the mind that most people would call the subconscious. I can see the decision-making process and I make decisions by consciously "clicking" on one of several different choices that appear as pictures.
The orienting response may be the beginning of consciousness, because it provides the animal with flexibility of behavior and it no longer has to rely on reflexes or hard wired instinctual behavior. When a deer hears a sound it freezes and orients. During orientation it makes a decision to either run away or continue grazing. Consciousness developed in the phylogenetically old parts of the brain so it is likely that even simple animals have a simple consciousness. Conscious thinking in mammals and birds enable flexible problem solving behavior in a novel environment. Mammals and birds are also socially conscious. Consciousness may be a matter of degree as brain complexity increases.

Key words

Autism, visual thinking, consciousness, categories.

According to some philosophers and scientists, animals are not fully conscious because they lack the ability to think in language. There is no internal language dialog in their brains. Budiansky (1998) has reviewed many studies regarding animal cognition and he thinks animals are not truly conscious because they do not have language. He recognizes the cognitive abilities of animals but he states "consciousness is quite another matter, though, for whether or not languages causes consciousness, language is so intimately tied to consciousness that the two seem inseparable" (Budiansky 1998, p193). According to Budiansky, I would have to conclude that as an autistic person who does not think in language that I am not conscious. There are many language based abstract concepts that I do not understand. To understand a concept I have to form a visual image in my imagination.

Now I would like to explain how I think visually. For example, when I think about what I have to do today, I see pictures in my imagination of going to the supermarket, and other activities. Language narrates the pictures and 'videotapes' that play in my imagination. When I design equipment for the livestock industry, I can test run it in full motion virtual reality in my mind. More on my visual methods of thinking are covered in *Thinking in Pictures* (Grandin 1995).

Visualizing equipment is easy because there are no language based concepts that cannot be visualized. Concrete things are easy to understand. Some philosophical writing is impossible for me to understand. Some of the words used in books about consciousness are so language based and abstract that I do not understand them. Words like "percept" and "mentalizing" make no sense to me. Even though I do not understand some language-based concepts, I think I am truly conscious. I am a college professor and have designed equipment that is used by most of the large meat companies n the U.S. and Canada.

Concept Formation Without Language

Forming categories is the beginning of forming concepts. The ability to form categories occurs in both people and animals. I am now going to explain how it is possible to form concepts by thinking with pictures. When I was a child I had to learn the difference between cats and dogs. At first I categorized them by size. That no longer worked when the neighbors bought a Dachund. Visual thinking is very specific. There is no generalized concept of a dog. There are only pictures of specific dogs in my memory. To comprehend the dog concept I had to look at many different specific dog and cat pictures I had stored in my memory. I had to find a common feature that all dogs had and none of the cats had. All the dogs, regardless of size have the same type of nose. All cats have an auditory feature that none of the dogs have. They all meow. My thinking is sensory based and not language based. After I determined the visual feature that all the dogs had, I could now categorize cats from dogs. The cat and dog pictures in my memory could now be sorted into dog and cat categories.

Categories can also be formed visually for more abstract concepts such as good and bad. My family taught me about good and bad behavior in a very concrete manner. Breaking a window would be bad behavior for which I would be punished and helping with the housework or cleaning up my room was good behavior. As I acquired many life experiences I could catalogue the specific experiences into different categories. Good and bad are not abstract concepts. They are like pigeon holes in a post office where the pictures of good and bad events could be catalogued. The concepts of good and bad were only understood when I had many specific images of good and bad experiences to look at. Then I can look for common features which make good and bad fall into visually understandable categories. Bad behavior hurts other people or their property. Bad behavior such as teasing can also hurt another person's feelings. When kids teased me, I cried. When I think about these events I see pictures in my imagination like videos. The pictures can then be sorted.

I am able to form concepts and make generalizations with visual pictures. Animals are also able to do this. For example, a guide dog for the blind has to be able to recognize an intersection in a strange city. Guide dog trainers teach the dog to generalize intersections by training the dog on many different types of interactions. If the dog was trained only on intersections with traffic lights, it may not know what to do at an intersection with no lights. Low functioning non-verbal people with autism have the same problem with generalizing. If the nonverbal person with autism is taught only at home to not run across the street, he or she will obey the rule at home but not at grandma's house. To generalize the non-verbal person with autism has to be taught not to run across the street at many different places.

My visual thinking for forming concepts is more complex. I use visual thinking for understanding concepts that are more complex than road safety rules. To understand how different parts of the brain are connected, I think about how the structure of the brain circuits is like a big corporation. I have been in many large corporate offices so I have many pictures stored in my memory. Neural circuits between different specialized areas of the brain are like different departments in a corporation connected by phone and computer lines. When a brain is damaged, some of the communication lines between departments are cut. The Chief Executive of the brain is the frontal cortex and when the frontal cortex is damaged, the different brain departments no longer work together in a coordinated manner.

Do people who never learn language as a child have true consciousness? SCHALLER (1991) describes teaching American Sign Language to a Mexican farm worker of normal intelligence who was deaf since infancy. This person had no education as a child and worked as a migrant laborer. Like a person with autism, he learned nouns first and his thinking was totally concrete and visual. She then describes observing several deaf farm workers who had never learned ASL who communicated in elaborate pantomime. They acted out their experiences of being chased by the border patrol. They took turns acting out their stories in great detail.

After the farm worker learned ASL he still never fully understood why people went to church and what an illegal alien was. Understanding these things requires complex use of language. He visualized that the border patrol (green uniform man) chased him away from the land of plentiful food and good jobs back to his homeland of little food. He had no concept of unjust or just, he just wanted to figure out the rules and avoid the green men. He never really understood why the green man kept chasing him. He did not understand God but he understood that a green work permit card had great power to repel a green man and make him stop chasing him back to Mexico. He did not know why the card had its power. The fact that his work permit card was fake was beyond his comprehension. Is he conscious? I would say yes. This man had normal emotions, communicated with facial expressions, did a variety of jobs and understood that work such as picking apples, got him money that he could buy food with. When he was a child, he knew there was something important in the books the other children used in school but he had no understanding of what printed words were. When he finally acquired language he would use single words to summarize his life. The word "green" had great significance. The great forces in his life were green; there were green men, green cards and green cars with green men in them. Money and the crops he picked were also green. SCHALLER (1991) describes visiting a house where several other deaf farmer workers with no language lived. They had a special place for their collection of green cards and they treated them like "gold." They did not understand an abstract concept such as God but they understood the power of their little pieces of cardboard which enabled them to stay in the land of more food.

Thinking with the Subconscious

I finally figured out that what FREUD called the unconscious is the part of the mind that people with autism and animals think with. If one thinks without language one has to have sensory based thinking. I think in pictures, a dog may think in smells. Animals recognize other animals and people by voice. Even specific vehicles can be recognized. Sensory based thinking is true thinking. BUDIANSKY (1998) provides an excellent review of animal thinking and cognition but he thinks that language is required for full consciousness. Research is making it very clear that animals think (GRIFFIN 2001). I hypothesize that in normal humans, language based thinking blocks access to more detailed sensory based thinking. Perhaps language blocks access to the subconscious. Research with patients with frontal temporal lobe dementia, a form of Alzheimer's disease, indicates that as the disease destroys the frontal cortex and the language areas of the brain, talents in art and music emerge (MILLER et al. 1998; MILLER/CUMMINGS/BOONE 2000). Many of these patients had no previous interest in art or music. To use my corporate office analogy, removing language provides access to the "art department" and the "music department" of the brain. The chief executive offices in the top office building are removed along with the legal department. This may also explain savant skills in autism. Savant talents are described in TREFFERT (1989) and HERMELIN (2001).

Research with normal people has shown that it is possible to gain privileged access to primary brain areas. Disabling the frontal cortex with a magnetic field will cause normal people to gain some savant skills in drawing (FOX 2002). Privileged access is explained further in (SNYDER/MITCHELL 1999).

I am able to use visual thinking in a more symbolic way to understand concepts in neuroscience. Brain scan research in autism has shown that autistics have direct access to the picture department in the brain. They excel at the hidden figures test. When they are put in a brain scanner while doing this test, only the visual part of the brain is activated (RING et al. 1999). In a normal person, many parts of the brain in addition to the visual area are activated. In my lectures, I often show two picture slides to symbolize this research. The brain scan of the autistic person is like a bright little cabin in a snowy dark wilderness and the brain scan of a normal person is like a bunch of lamps in a lamp store. In the normal person, it is difficult to tell which brain area is activated specifically for the hidden figure test. The activity in the visual part of the brain is lost among all the other activated areas. Maybe this is why the normal person does more poorly on certain visual tasks. The other parts of the brain may interfere with visual perception.

My Own Experience Thinking with the Subconscious

Below I describe how I may be directly accessing what most people would call the subconscious. In the following description of how I avoided a car accident, I explain how I used thinking in pictures to make conscious decisions. This example illustrates a level of consciousness that may be in some ways similar to consciousness in higher mammals. The near-accident occurred in fairly light traffic on a sunny day while I was driving to the airport on Interstate Highway 25. Cruising along at 70 mph in the southbound lane, I suddenly saw a huge bull elk running full speed across the northbound lanes. I knew I had to react quickly to avoid hitting him. Instantly, three pictures appeared in my mind. Each picture represented the end result of an option available to me. The first pictures was of a car rear ending my car. I knew from experience that slamming on the brakes could cause this. The next picture was the elk smashing through my windshield. From my understanding of animal behavior, I knew that swerving or any sudden sideways movement of my car might cause the elk to stop or slow down. The third picture was the elk passing harmlessly in front of my car. In this picture I saw what would happen if I gently applied the brakes to slow down. These pictures were like the picture menus one can click on an Internet web page. They appeared in my mind one at a time, but all within one second. This was enough time for me to selectively compare the options and choose the slow down gradually picture. I immediately calculated the elk's trajectory and speed coming across the highway, and my speed and position in the southbound lane and began to slowly apply the brakes. This choice prevented me from being rear ended, or having the elk crash through my windshield. The conscious choice was a visual process without the use of internal verbal dialog.

At the moment I became aware of the elk crossing the northbound lane, I resisted the urge to make a panic response and slam on the brakes. In just seconds, I evaluated the three pictures in my mind. To use computer jargon, I conducted a basic cost–benefit analysis of the options. After running a quick video like simulation of the elk passing harmlessly in front of my car, I simply clicked a mental mouse on the "slowing down gradually" picture. I made a conscious choice from visual simulations played in my mind. In another mishap on the highway, my ability to make a conscious choice was overridden by sudden panic. I was driving along a section of straight level highway on an icy night when a sudden gust of wind caused the car to skid. In this situation, I did not have time to make a conscious decision. Conscious behavior can only occur when there is time to think, whereas instincts, reflexes and simple conditioned responses take over when there is no time to think. For example, a grazing animal suddenly being attacked or chased by a lion relies on instincts and reflexes. These behaviors may not be completely conscious. However, when an approaching predator is far away, an animal has time to decide on the best evasive action. When I hit the patch of ice, reflexes took over and I lost the ability to make an appropriate response. No option pictures appeared in my mind which could be used for making a decision. Reflexively, I began swearing uncontrollably and jerking the wheel in the wrong direction as I was skidding off the highway. I had no time to recall what I had learned about steering into a skid. My car ended up on the median strip and fortunately, my vehicle and I were undamaged.

Some people question why I had three visual choices instead of just one. I think language covers up seeing the choices. This is due to my visual associative way of thinking. In everything I do, I see different choices as pictures on a computer monitor in my imagination. My thinking is not linear. I have learned by interviewing highly verbal thinkers that their thoughts are in language and they do not consciously see choices. Language may be another layer of thinking which covers up the visual pictures. I have no purely abstract thoughts. I only have pictures.

The "autistic type" of consciousness I used in both near accidents may be in some ways similar to conscious processes some animals use when they encounter danger. In both animals and people, conscious processes may have evolved as mechanisms for both avoiding danger and finding food. In other words, consciousness evolved as a means of allowing higher mammals to perform intelligent, adaptive responses to challenges in their environment. Rather than always relying on reflexes, simple conditioned responses, or hard-wired instinctual behavior patterns, consciousness allows animals to make choices between several different options. Although consciousness is important, instinctive, reflexive and simple learned behaviors are also important. The instinctive killing bite to the throat used by most predators, the reflexive response of a horse kicking at a predator on its heels, or the conditioned response of learning to avoid places that are full of predators, all evolved as mechanisms used for survival and may not require consciousness. Even insects can learn a simple conditioned response. The questions of whether non-human animals have consciousness depends on what we mean by consciousness. Animals are probably conscious if you can agree that consciousness without language is possible.

Orienting Response is the Beginning of Consciousness

On Thursdays, the garbage truck picks up trash in the neighborhood next to where Mark stables his horses. The moment the back up alarm of the truck sounds, all the horses turn and orient towards the sound. Like soldiers at attention, all the horses aligned their eyes, ears, head and body in the same direction. The orienting response is accompanied by increased heart rate, respiration and blood pressure. When animals orient they switch from unconscious behavior to conscious. Both animals and people orient towards novel sounds. In the wild, animals orient and freeze when they hear or see something that might be dangerous. A deer that hears the rustling sound in the bushes instantly freezes and turns both its eyes and ears towards the sound. A deer will turn and face the noise before it flees. The orienting response provides time for the animal's brain to make a conscious decision instead of just acting on reflexes and instinct. During the orienting response, the deer can decide to either flee or continue grazing. When I avoided the elk on the highway, I had time to make a conscious choice. But, when I skidded on the ice, there was not enough time to make a conscious choice.

LIBET's research suggests that during the orienting phase the brain can consciously veto a response. In my own case, the first picture that popped into my imagination was the consequences of a panic response. The third and best response, which was the last picture to occur, required suppression of the reflex panic response. To put it in more philosophical terms, the brain does not allow free will but it definitely gives you powerful abilities to veto certain responses. To exercise the veto power, there must be time to look at different possible responses. Research by Benjamin LIBET at the University of California has shown that the brain takes longer to process conscious awareness of a stimulus compared to an unconscious reaction to it. Up to half a second is required for full conscious awareness to occur after a stimulus is applied to the brain (NØRRETRANDERS 1991). If you touch a hot stove, an unconscious reflex controlled by your spinal cord has already pulled your hand away before you feel the pain. Conscious processing of incoming information takes more time than a simple response governed by a reflex. A zebra kicking at a lion is probably relying on reflexes, but a zebra that hears a far away sound which may signal danger has time to weigh his escape options.

Levels of Consciousness

Brains become more complex when the phylogenetic tree is ascended. The brain expands and more and more areas are interconnected. Consciousness becomes more complex. Being a visual thinker, I want to look at concrete things I understand such as comparing nervous system complexity between different species. I agree with William JAMES, the father of psychology, who stated in 1891, "consciousness grows more complex and intense the higher we rise in the animal kingdom" (JAMES 1891, p141). LEDOUX (1996) maintains that conscious occurred in the animal kingdom when the cortex expanded and it allows animals to relate several different things at once. Both DAWKINS (1993) and GRIFFIN (2001) agree that conscious behavior occurs when animals can adapt their behavior and solve problems under novel conditions. To be conscious requires flexible behavior.

I have to conclude that there are some higher language based conscious experiences that I do not have. There are books on philosophy that make no sense and the world of algebra is impossible for me to understand. A problem with discussing consciousness is that some of the discussion goes into abstract language based concepts I simply do not understand. To define consciousness one must first define the word 'conscious.' I prefer definitions based on nervous system complexity. I think it is wrong to say that the term consciousness only applies to humans with language. DAMASIO (1999) states that consciousness resides in the phylogenetically old part of the brain. When a person gets Alzheimer's, consciousness is one of the things they lose last. They are fully aware of losing their other abilities.

Visual thinking is also phylogenetically old. A recent review by COLLETT/COLLETT (2002) in *Nature Review in Neuroscience*, shows that insects use visual thinking to navigate. Very small brains can store visual information and compare an image that is observed with an image in memory. Do I think individual bees and ants are conscious? My answer is *no*. Electronic circuits can be made to mimic insect behavior. So as a person with autism, what is my definition of consciousness? For me, consciousness and being able to think are the same thing. Thinking is being able to solve problems in a novel situation through the use of previously learned information (GRIFFIN 2001; DAWKINS 1993). It is not reflexes, instinct or simple operant or classical conditioning. There is a vast scientific literature on animal cognition.

Consciousness Fragments in Abnormal Brains

Behavior becomes more complex in more complex brains. Damage to the frontal cortex in both animals and people reduces the flexibility and complexity of behavior (KOLB 1990; FREEMAN/WATTS 1950). Flexible behavior requires a brain that can associate information from many of its parts. Research in autism by Nancy MINSHEW and her colleagues at Carnegie Mellon University show that in autistic brain, the different parts of the brain are less interconnected (MINSHEW 2002, personal communication). Are they less conscious than a normal person? I will let the language-based philosophers decide.

Many people with autism, myself included, have problems with multitasking. Autistics also have difficulty hearing and seeing at the same time. Donna WILLIAMS, a verbal writer with autism, explains how her senses fragment (WILLIAMS 1988). She writes "ongoing conscious awareness is a luxury that an overloaded nervous system cannot afford" (WILLIAMS 1998, p239). She has difficulty integrating her emotions and thoughts together. Emotions are disconnected from thoughts. In my own career, I can play video in my mind of past events with little or no emotion. From my own experiences and from reading Donna WILLIAM's writing, it has become obvious that all the different subsystems are not working together. To use my corporate office analogy, Nancy MINSHAW's research shows that the autistic brain would be like an office where both internal phone lines and lines to the outside world are malfunctioning. Communication between departments inside the office building are hindered and sensory circuits that would be analogous to phone lines from the outside world also malfunctioning. There are fewer phone lines and they are often overloaded and are full of static. It is my opinion that when Donna's sensory system overloads and malfunctions, she has impaired consciousness. Thinking consciously works fine for me, but my emotional consciousness is impaired because I do not fully integrate thought and emotion. A brain scan test showed that an area in my frontal cortex is missing a circuit to the amydala. The amydala is the brain's emotion center.

Compared to some animals, my emotional and social consciousness is impaired. I was over 50 years old when I learned that people communicate emotion with subtle eye movements. I read about them in a book entitled, *Mind Blindness* (BARON-COHEN 1995). GRIFFIN (2001) states that animals are able to convey some of their thoughts to other animals. Prairie dogs have a complex communication system of tones for communicating to other prairie dogs information about predators (SLOBODCHIKOFF 2002). They have specific tones and calls that they invent for warning of specific individual predators such as "coyote that stalks", or "coyote that lies and waits". The calls convey both emotional urgency and a description of the predator.

They have specific tones and calls that they invent for warning of specific individual predators such as "coyote that stalks" or coyote that lies and waits. These calls convey both emotional urgency and a description of the predator.

Lack of Sensory Integration and Consciousness

I think that the fragmented consciousness that a person with severe autism has may be similar to some lower animals. The lower one descends on the phylogenetic tree, the more difficulty an animal has with association across the senses of vision, hearing or touch. Maybe some animals are only conscious on one sense and it is all reflexes and instincts and simple learning in others.

Moving up the evolutionary ladder from insects, many biological scientists agree that mammals and birds have primary consciousness because they can process simultaneous stimuli and they have an internal representation of their experiences. Birds are capable of problem solving under novel conditions. Sverre SJÖLANDER (1997) states that a snake may not be conscious because it does not have a centralized representation of its prey. It seems to live in a world where a mouse is many different things. SJÖLANDER explains that striking the mouse is controlled by vision; following the mouse after striking is controlled by smell; and swallowing the mouse is controlled strictly by touch. There is no integration of information from all the senses. Each sensory channel operates independently of the others. When a snake has a mouse held in its coils, it may still search for the mouse as if "the information from its body which is holding the prey did not exist" (SJÖLANDER 1997, p597). It appears that the snake has no ability to transfer information between sensory channels (STEIN/MEREDITH 1994; SJÖLANDER 1995). SJÖLANDER (1997) further explains that a snake has no ability to anticipate that a mouse running behind a rock will re-appear. Cats and other predatory mammals are able to anticipate that the prey will reappear. According to SJÖLANDER, snakes are not conscious. Using this definition of consciousness, then an autistic person experiencing severe sensory overload is also not conscious. Sensory overload causes them to lose the ability to integrate input from all the senses.

Lower animals such as reptiles and insects have great skills that resemble the skills of an autistic savant. The autistic savant can do great feats in art, music or mathematical computation but he is incapable of many ordinary everyday skills such as cooking or balancing a checkbook. Simple social interactions have to be learned by rote. Ants and bees have a great ability to store and use visual images (COLLINS/COLLINS 2002) but there is no evidence that they can do flexible problem solving. There is simply not enough association circuits in their simple nervous system.

Maybe there are levels of consciousness that can be correlated to how the nervous system is wired. Perhaps bees are visually conscious but have no other consciousness?

Conclusions

To have a sensible discussion on consciousness one must first define consciousness. Griffen (2001) also discusses the problem of defining consciousness. To me, consciousness implies flexibility of behavior under novel conditions. Maybe there are different levels of consciousness.

1. Consciousness within one sense. Example: ants using stored visual images to navigate. I would classify ants as not conscious unless they can solve a problem under novel conditions.

2. Consciousness of aversive stimuli that cause fear possibly in only one sense. I will speculate that fish operate at this level.

3. Consciousness where information from all the senses can be integrated with emotion and the emotions are more complex than just fear. I speculate that dogs and birds operate at this level.

4. Consciousness in people who have normal intellect but the emotions are not fully integrated due to missing connections between the frontal cortex and the limbic system. I put myself in this category. Some autistics will lose consciousness in one or more sensory channels.

5. Consciousness where all thinking is in language. The thing that is interesting is that I have observed from interviewing many people that people who think totally in language are poor in art. Their minds are language specialists. People who are good at art or engineering seem to have more specialized minds that are poor at social emotional consciousness but good at more specialized uses of the mind.

To conclude, maybe consciousness is like GARDNER's (1983) different types of thinking. Consciousness may vary in both type and complexity. To have consciousness there probably has to be a certain minimum amount of circuits to associate information from different parts of the brain. Maybe there is fear consciousness, pain consciousness, visual consciousness, language abstract consciousness, smell consciousness and many other combinations and types of consciousness.

Author's address

Temple Grandin, Department of Animal Sciences, Colorado State University, Fort Collins, CO 80523-1171, USA.
Email: cmiller@ceres.agsci.colostate.edu

References

Baron-Cohen, S. (1995) Mindblindness: An essay on autism and theory of mind. MIT Press: Cambridge MA.

Budiansky, S. (1998) If a lion could talk: Animal intelligence and the evolution of consciousness. The Free Press: New York NY.

Collett, T. S./Collett, M. (2002) Memory use in insect navigation. Nature Reviews Neuroscience 3:542–552.

Damasio, A. (1999) The feeling of what happens: Body emotion and the making of consciousness. Harcourt Brace: New York NY.

Dawkins, M. S. (1993) The search for animal consciousness. W. H. Freeman: New York NY.

Fox, D. S. (2002) The inner avant. Discover Magazine, February:44–49.

Freeman, W./Watts, W. (1950) Psychosurgery (2nd Edition). C. C. Thomas: Springfield.

Gardner, H. (1983) Frames of mind: The theory of multiple intelligences. Basic Books: New York NY.

Grandin, T. (1995) Thinking in pictures and other reports from my life with autism. Double Day: New York NY. Now published by Vintage Press, division of Random House.

Griffin, D. R. (2001) Animal minds: Beyond cognition to consciousness. University of Chicago Press: Chicago IL.

James, W. (1891) The principles of psychology, MacMillan: London. Reprinted in 1981 by Harvard University Press, Cambridge MA.

Kolb, B. (1990) Prefrontal cortex. In: Kolb, B./Deer, R. C. (eds) The cerebral cortex of the rat, Cambridge MA: MIT Press.

Hermelin, B. (2001) Bright splinters of the mind: A personal story of research with autistic savants. Jessica Kingsley Publishers: London UK.

LeDoux, J. (1996) The emotional brain. Simon and Schuster: New York NY.

Miller, B. L./Boone, K./Cummings, J. L./Read, S. L./Mishkin, F. (2000) Functional correlates of musical and visual ability in frontal temporal dementia. British Journal of Psychiatry, 174:458–463.

Miller, B. L./Cummings, J. L./Boone, K. (1998) Emergence of artistic talent in frontal temporal lobe dementia. Neurology 51:978–981.

Nørretranders, T. (1991) The user illusion. Cutting conscious down to size. New York NY: Viking. Translated to English, 1998.

Schaller, S. (1991) A man without words. University of California Press: Berkeley CA.

Sjölander, S. (1995) Some cognitive breakthroughs in the evolution of cognition and consciousness and their impact on the biology of languge. Evolution & Cognition 1:3–11.

Sjölander, S. (1997) On the evolution of reality: Some biological prerequisites and evolutionary stages. Journal of Theoretical Biology 187:595–600.

Slobodchikoff, C. N. (2002) Cognition and communication in prairie dogs. In: Bekoff, M./Allen, C./Burghardt, G. (eds) The cognitive animal. MIT Press: Cambridge MA.

Stein, B. E./Meredith, M. A. (1994) The Merging of the Senses. MIT Press: Cambridge MA.

Snyder, A. W./Mitchell, D. J. (1999) Is integer arithmetic fundamental to mental processing? The minds secret arithmetic. Proceeding Royal Society of London Biology 266:587–592.

Ring, H. A./Baron-Cohen, S./Wheelright, S./Williams, S. C./ Brammer, M./Andrew, C./Bullmore, E. T. (1999) Cerebral correlates of preserved cognitive skills in autism: A functional MRI study of embedded figures and task performance. Brain 122:1305–1315.

Treffert, D. A. (1989) Extraordinary people: Understanding the savant syndrome. Harper and Row: New York.

Williams, D. (1988) Autism: An inside out approach. Jessica Kingsley: London UK.

Filip Geerardyn/Gertrudis Van de Vijver/Veroniek Knockaert/
Ariane Bazan/David Van Bunder

"How Do I Know What I Think Till I Hear What I Say"

On the Emergence of Consciousness Between the Biological and the Social

Abstract

The commonsense exclamation "How do I know what I think till I hear what I say" is taken to reflect what happens in a clinical psychoanalytic session when someone becomes conscious of something that was previously unconscious. Moreover, it is in agreement with FREUD's theory of this phenomenon, more especially his theory on "derived consciousness" (NATSOULAS 1985). It is argued (i) that derived consciousness, although rooted in biological and neurological mechanisms, cannot be reduced to them, because consciousness emerges between the levels of the biological on the one hand and the social or linguistic on the other hand; (ii) that FREUD's theory contains no dualistic conception of the mind/body problem but is to be qualified as a non-reductionist materialistic theory. It is suggested, finally, that the notions of emergence and signifier can contribute to bridge the gap between neuroscience and psychoanalysis in that they inspire new experimental research paradigms that have some external validity with respect to the clinical reality of psychoanalysis on the one hand, and at the same time are acceptable within the methodological traditions of neuroscience.

Key words

Consciousness, unconscious, evolution, emergence, signifier.

Introduction

Recently, in their target paper on "The Unconscious Homunculus", Francis CRICK and Christof KOCH (2000a, p7) focussed on the relation between visual perception and consciousness and referred to the commonsense exclamation "How do I know what I think till I hear what I say" in support of their view that consciousness involves perception, and as such requires a contribution from motility. A bit odd however is their somewhat irritated conclusion of their response to the comments they received:

"This leads to our final comment, which may not be a welcome one. We are disappointed that (with the exception of LIBET and SCHALL) no one else made any suggestions of any experimental test of a neurobiological nature that might advance the subject. If psychoanalysis and neuroscience are to interact effectively there must be more emphasis on possible experiments, especially neuroscientific ones, and less time devoted to describing, ad nauseam, what people thought in the past. One of us proposes that, in the future, any reference to 'FREUD' should be barred from this journal [Neuro-Psychoanalysis!], at least for the next 10 years, though, of course, one could hardly object if authors read his work in secret" (CRICK/KOCH 2000b, p58).

CRICK and KOCH are with no doubt right in stressing the importance of experimental research. But however important fundamental research of the visual cortex in monkeys is in contributing to a solution of the mind/body problem, we imagine that it can hardly excite psychoanalysts involved in clinical practice. If the aim is to bridge the existing gap between actual neuroscience and psychoanalysis, any new experimental paradigm should keep an eye on both neuroscience and psychoanalytic practice.[1]

"How do I know what I think till I hear what I say"

The commonsense exclamation "How do I know what I think till I hear what I say" refers to the familiar phenomenon in which, in addressing oneself to another person or to an audience, one suddenly hears oneself saying something that one did not really intend to say, but which, once said, is recognized as an expression of ones own thought. The exclamation can thus be translated as follows: "Sometimes I am not conscious of my own thoughts, in any consciously anticipated sense, unless to my own surprise, I acknowledge them in the words I have just spoken". In that sense, speech is a kind of facilitator for the becoming conscious of some thought that in one way or another was already there. One might think here of what is at stake in the technique of brainstorming, but psychoanalysts refer us to the paradigm of the FREUDIAN slip of the tongue in which a repressed intention, desire or thought suddenly "emerges" or re-emerges. It was this facilitating function of speech for the process of the becoming-conscious that inspired FREUD to introduce the fundamental rule of free association in his therapeutic device. Initially, that rule focussed on remembering and memory and was meant to make the patient conscious again of some thoughts that already as such existed (cf. e.g. FREUD 1953c).

Very soon however, in his famous letter of December 6, 1896, to his Berlin colleague Wilhelm FLIESS, a much more sophisticated conception of the process of the becoming-conscious transpires. There he (FREUD 1966b) writes, "A failure of translation [or wording or symbolising of material memory traces]—this is what is known clinically as 'repression'" (p235). Implied in this definition is his stratified conception of the psychical apparatus according to which any memory traces of past sensory impressions can only attain a subsequent level when they are "transcribed" or "translated". As a consequence, FREUD (1966a) could only conceive his psychoanalytic treatment as the accomplishment of "a psychical act, which did not take place at the time", (p300), that is of the psychical act of symbolisation through which repression is undone. Hence, for FREUD speech not just facilitates the process of becoming-conscious of ones own thoughts, that are already there, but it also involves a *creative* aspect.

The latter aspect has been very well rendered in the more recent LACANIAN distinction between "empty" and "full" speech, between speech that has been "constituted" beforehand, and speech that is actual "constitutive" (cf. KNOCKAERT 2001, p100; LACAN 1977, pp40, 48–49). It is definitely the latter that is aimed at in psychoanalysis in that it aims at "translating" the material memory traces into signifiers[2], as a consequence of which, through the act of speaking, one becomes conscious of ones own thoughts or intentions. In a psychoanalytical sense, the expression "How do I know what I think, till I hear what I say" can be adequately translated by the LACANIAN saying that the unconscious can only emerge in speech.

Let us now take a closer look at the FREUDIAN theory of the mental apparatus and its functioning, a theory that he first elaborated in his so-called "Project for a scientific psychology" (FREUD 1966a), his "psychology for neurologists" (FREUD 1985, p127). What is at stake in his "Project" is nothing less than the ambitious endeavour (1) to provide us with a theoretical understanding of the dynamics observed in clinical practice—i.e. the fact that the process of the becoming-conscious requires speech, or, to put it in other words, the efficiency of the rule of free association (GEERARDYN 1997, pp242–247); (2) to embed these dynamics in a neurobiological theory on the development of the mental apparatus.[3] Here we will only highlight two aspects of that theory, i.e. FREUD's description of the contributions of speech motility on the one hand, and of the acoustic element in the process of the becoming-conscious on the other hand.

The Contribution of Speech: "…What I *Say*"

In his "Project" which aims at providing "a psychology that shall be a natural science: that is, to represent psychical processes as quantitatively determinate states of specifiable material particles" (1966a, p295), FREUD starts from the idea of a single neuron N that is governed by the constancy principle in that it tends to keep the quantity within it as low as possible. Any rising of its quantity, caused externally or internally, gives rise to unpleasure. In that case, the organism will get rid of the surplus of quantity by motor discharge.

According to FREUD, a psychology worth of any consideration should be able to explain the important mental functions of memory and consciousness. In order to do so, he functionally differentiates the ensemble of neurones and attributes the functions to distinct neurological systems or classes of neurones of the mental apparatus. Memory belongs to the system while consciousness belongs to the system. Any external perception or experience has different consequences for these systems. On the one hand, the quantitative aspect of the experience endurably changes the memory system, in that any passage of quantity will facilitate the "contact-barriers" (or synapses) between the neurones of the memory system. That is, in the mental apparatus, memory is laid down in the contact barriers, or, more precisely, is "represented by the differences in the facilitations between the neurones" (ibid., p300). On the other hand, the second characteristic of the same experience, its period or quality, is transferred to the system where it gives rise to consciousness (ibid., pp308–309).

The result of any primary experience will be that in the memory system a residue of that experience is left that will influence all subsequent processes. For, in accordance with the principle of facilitation, any second experience leads to a renewed activation of the memory of the first experience. This implies a danger for the organism: when the memory of a primary experience of satisfaction is freshly activated, then it is of the utmost importance to know whether the object of satisfaction is actually present in the external world. That is, the ego needs a criterion or an indication in order to be able to distinguish between a perception and a reactivated memory.

This criterion is termed the "indication of reality" (ibid., pp325–326). For every external perception a qualitative excitation occurs in the consciousness system, resulting in a motor discharge of which the memory system is informed. By contrast, when only a memory is recathected, there will be of course no such indication of reality.

For the primitive ego, which only seeks to repeat the primary experience of satisfaction, this indication of reality functions as a sign to pay attention to what is out there, to the object of satisfaction. Here the attention mechanism is introduced as a mere feedback principle that serves the pleasure ego, to which now three possibilities are open: (1) In case of a complete correspondence between perception and memory, the experience of satisfaction can be repeated without danger; (2) In case the correspondence is only partial, then the mnemic image will be cathected with quantity from both sides, i.e. from the external world and from the ego. This results in what FREUD calls a thought process, which in fact aims at installing the identity, as fast as possible, between perception and memory; (3) Perception and memory do not correspond at all. In this case, the mechanism of attention is of no use to the pleasure ego, unless *language* comes into the picture.

Indeed, FREUD assumes that it is always in the biological interest of the ego to cathect perception with additional quantity, for one never knows, one might eventually find back the primary object of satisfaction. But where then does the additional quantity come from, if not from any indication of reality, which results, as we remember, from a motor discharge? "This purpose is fulfilled by speech-associations. This consists in the linking of neurones with neurones which serve sound-presentations and themselves have the closest association with motor speech-images" (ibid., p365).

This speech offers an enormous advantage to the ego in that it (1) results in a cognising of the world—indeed, "How do I consciously know what I perceive till I hear what I say"; (2) And at the same time ensures a gradual discharge of quantity. More importantly, this model implies that for FREUD thought-processes as such have no sensory quality and are therefore, by definition, unconscious. In human beings, however, the latter may acquire or "derive" (NATSOULAS 1984, p227) quality from qualitative verbal memories.

The Contribution of the Acoustical Element: "… Till I *Hear* …"

From a developmental viewpoint however, FREUD pays more attention to the acoustical element.[4] For the human infant biological conditions are such that it is born as it were prematurely, as an extremely helpless creature (FREUD 1966a, p318). In case large quantities of stimulation, emanating from the external world affect the child, it will attempt to rid itself of these quantities, at first reflexively, through its body movements or motor innervations. However, quantities are also released from within the body. These endogenous quantities may be small but they are released constantly, as is the case when the child becomes hungry or thirsty and thus finds itself in what FREUD calls conditions of "Lebensnot" or "the exigencies of life" (ibid., p297). The child also tries to discharge these quantities but it is obvious that here a "specific action" needs to be performed in the external world, which cannot be done by the child. It is simply unable to make use of the constant accumulation of endogenous quantities in any efficient manner and can only rely on the "Nebenmensch", its fellow human being.

In FREUD's view, the sequence of events can be presented as follows: the accumulated quantities are first discharged through motor innervations in the crying of the child. Then a fellow human being, e.g. its mother, "interprets" this crying as the child's being hungry or thirsty and feeds it, which brings about satisfaction or a discharge of the excitation:

"In this way this path of discharge acquires a secondary function of the highest importance, that of *communication* ["Verständigung"], and the initial helplessness of human beings is the primal source of all *moral motives"* (ibid., p .318).

This narrative or theoretical fiction is about what FREUD calls the *primary experience of satisfaction* and highlights the importance of two conditions (1) the biological condition of the prematurity of the human child; (2) the social or cultural condition that when the child is born *language is already there*. As a consequence, from a developmental point of view, we can easily understand that FREUD privileged the sound-image or acoustic aspect of language. For the latter is the one human beings are first confronted with (KNOCKAERT et al. 2002).

Clinical: "... Till I Hear What I Say"

From a clinical psychoanalytical point of view, both the acoustic and motor images are important. What FREUD was interested in was of course a psychological theory that could explain the empirical phenomena he met in his practice. Amongst other the important fact that his patients only became conscious of unconscious desires through speech.

But what is it then for an unconscious process to become conscious? For FREUD, this does not mean that the unconscious process as such becomes a conscious one. It only means that it becomes associated with a corresponding "intrinsically" (NATSOULAS 1984, p198) conscious process such as a perception. Unconscious processes "urge" towards consciousness and try to gain access to the perception-consciousness system, to an *actual* process that occurs there; it is a matter of their becoming associated with the residues or memory images of perceptions of words (FREUD 1957b, p202). The latter have as yet no quality and therefore are descriptively unconscious and must be situated, in FREUD's topographical model, in the pre-conscious, but they will in the end enable the process of the becoming-conscious through *actual speech*. And it is only then that we arrive at the situation described in the commonsense exclamation "How do I know what I (unconsciously) think till I hear what I say?" Indeed, nothing of the unconscious can be known unless it passes by consciousness. From a clinical psychoanalytical point of view this process involves an important aspect: it is only in hearing oneself say something that one can acknowledge one's thought as one's own. For as long as one does not acknowledge this thought through hearing oneself expressing it, it will only transpire unconsciously in one's symptoms and other formations of the unconscious. More theoretically, it is only through speech that an unconscious thought acquires a subjective side, which is, according to FREUD, characteristic of conscious processes (FREUD 1966a, p311; NATSOULAS 1985, pp210–215)[5].

Evolution

From an evolutionary point of view, FREUD's model implies that (1) primitive man from the moment articulate speech became biologically possible, in the very act of articulating created himself an extra *feedback-loop* and as such, an extra sense, a *judging sense* that very closely associated the mnemic image of the motor speech images of what he was articulating with the mnemic images of what he heard himself say (as well as with any other actual perceptual images, stemming from the activation of the other senses); (2) although primitive man, being a complex biological organism, no doubt was intentionally directed towards the external world, he by no means was anticipating any meaning when he tried to grasp the world with his grumbling. For any possible meaning, in its psychological sense, could only result or "emerge" from his grumbling being differentiated, that is from articulation and association; (3) the commonsense expression "How do I know what I think till I hear what I say" of course remains true for primitive man in that he only started to know what he thought after he heard what he was saying.

Further, this model implies that also from an evolutionary point of view, it will be important to define the precise nature of articulate language. At least one aspect becomes clear from what we have said so far: articulation or the introduction of differences implies the introduction of a special temporality—man does not express what he means, but he means what he expresses—the temporality of what FREUD called "deferred action" or "Nachträglichkeit". This means that although articulation itself proceeds chronologically, or is as it were a movement forwards in time, the meaning of what is articulated only emerges at the end or in the associations of the articulated differences. To be retained here is the fact that this temporality seems to have a biological ground in the very construction of the speech apparatus and its interaction with the auditory apparatus: the hearing comes *after* the saying.

Emergence

So far, we have argued that the way in which the FREUDIAN unconscious becomes conscious through judgement or speech is a phenomenon that is adequately described in the commonsense exclamation "How do I know what I think till I hear what I say". It was this phenomenon that FREUD put in the centre of his clinical practice as well as of his theory. Starting from the idea that many of the symptoms and the sufferings of his patients were unconsciously motivated, i.e. that they should be considered as the non-acknowledged expressions of unconscious motives or desires, he developed a therapeutic device that facilitated acknowledgement of these desires, that is the becoming-conscious of them. In his presentation of the psychical apparatus, written down in his "Project" and refined and elaborated in his metapsychological writings he tried to understand this phenomenon, amongst other things, theoretically.

But what else can be said about this theory?

1. The scope of his theory was far broader than psychopathology and the clinical phenomenon he departed from, i.e. the phenomenon of the becoming-conscious of the unconscious. Evidence for this can be found in his explicit ambition, transpiring from the very start in his "Project", to explain normal psychical processes as well (1966a, pp360–387). His theory concerns conscious processes and about consciousness as much as it is about unconscious processes and about the unconscious—logically enough, given the fact that the unconscious can only be known through consciousness.[6]

2. Here we have read FREUD's theory as being a theory on how the unconscious becomes conscious, or on how the unconscious "emerges" and at the same time is consciously acknowledged through speech. But this kind of consciousness which he considered being the highest accomplishment of evolution and which accounts for the superiority of man over other animals (FREUD 1953b, p617) is conditional. FREUD's theory reckons with the biological conditions, the most important of which must be the biological conditions of articulate speech. As we have seen from the developmental point of view that is contained in FREUD's model, there is however also a social condition in that language precedes the individual. Thus his theory is not just a theory on how the unconscious appears or "emerges" in the descriptive meaning of the verb "to emerge". We argue that his theory in fact combines materialism with a robustly causal version of emergence. For in FREUD's view unconscious processes cannot but urge to "gain access" to the pre-conscious word-presentations as residues of human discourse, of what previously has been heard. As LACAN (1974, p16) would have it: "l'inconscient, ça parle" which means "the unconscious speaks" or even "the FREUDIAN Id speaks". The Id manifests itself through the symbolic order, which precisely precedes the subject[7]. This became evident for FREUD from the very start of his clinical experience that witnessed of the fact that symptoms and dreams could be unmasked as being non-adequate, i.e. non-acknowledged transcriptions or translations of the unconscious.

In this way, FREUD's theory defines the genuine human mental or the psychical processes that involve language, as emerging in-between the organizational levels of the biological conditions and the social condition (cf. VAN BUNDER 2001; VAN BUNDER et al. 2001). As a biological organism the human being at all times manifests a *biological compliance*, is intentionally directed towards the external world, towards the other as object… only to meet the *linguistic compliance* of the *signifier.*

Conclusion

For a long time academic psychology suppressed the scientific study of consciousness in considering the latter as being a mere epiphenomenon of the biological or neurological. Quite similary it can be argued that psychoanalysis has for a long time suppressed the biological nature of FREUD's theory. This "suppression" can be explained by (1) the horror of both an ontological reduction of mental mechanisms to biological and neurological mechanisms and of the epistemological reduction of psychoanalytic theory to biology and neurology; (2) the fact that the latter reduction does not make any difference for their clinical practice. However, when we bear in mind what FREUD said about

Author's address

Filip Geerardyn, Research Unit in Neuro-Psychoanalysis at the Department for Psychoanalysis and Clinical Consulting at Ghent University, H. Dunantlaan 2, B-9000 Ghent, Belgium
Email: filip.geerardyn@rug.ac.be

the acoustical word-image, further elaborated by LACAN in his conception of the material signifier—i.e. its being a physical object—, then psychoanalytic theory appears as a theory that accounts for the causal power of the signifier and of speech. In this way, LACAN's idea of the materiality of the signifier demystifies psychoanalysis and can help to bridge the gap with science. For being a concrete physical object that affects mental and even neurological and biological functioning, the signifier can perfectly serve as a paradigm for scientific experimental research (cf. BAZAN et al. 2002).

Notes

1 Although "there are numerous features of FREUD's theories that reflect nineteenth-century misconceptions of the nervous system" (SMITH 2000, p39) and that his "ideas map rather poorly onto what we know today of the primate brain" (CRICK/KOCH 2000b, p51), FREUD consequently elaborated his theory while keeping an eye on both his clinical practice and on the neuroscience of his day.

2 LACAN borrowed his notion of the signifier from the work of Ferdinand DE SAUSSURE. For the latter the signifier is the phonological element of the sign, i.e. the mental acoustic image—cf. EVANS (1996, p186), DE SAUSSURE (1916, pp 66–67).

3 Although FREUD never published his "Project"—it was contained as a manuscript in his correspondence with FLIESS—, he never abandoned it but further developed this theory throughout the rest of his writings, in his Interpretation of dreams (1953b), his metapsychological papers on "Repression" (1957a) and on "The unconscious" (1957b), in his essay on The Ego and the Id (1961a), in his "Note on the mystic writing-pad" (1961b), in his "Outline of psycho-analysis" (1968) and many other writings.

4 As was the case in his book "On Aphasia", published four years earlier than the "Project", where he had privileged the sound image in his definition of meaning according to which a word acquires its meaning through the association of the sound image with the object presentation (FREUD 1953a, p.77).

5 According to FREUD, all conscious thought processes involve an activation of motor-representations of speech. In the thinking adult, however, this activation is small and manifests itself in silent or sub-vocal speech.

6 No wonder then that NATSOULAS (1984, pp195–196), in his comprehensive study on FREUDIAN consciousness argued that academic psychology, now that the "suppression of the scientific study of consciousness" has come to an end, might very well profit from taking a closer look at FREUD's theory of consciousness.

7 Here we follow VAN GULICK (2001, pp19–20) who states that the challenge to those who wish to combine materialism with a causal version of emergence "is to find a way in which higher-order properties can be causally significant without violating the basic causal laws that operate at lower physical levels. (…) One possible solution would focus on the respect in which higher-order patterns might involve the selective activation of lower-order causal powers".

References

Bazan, A./Geerardyn, F./Knockaert, V./Van de Vijver, G./Van Bunder, D. (submitted) Role of the phonology in the neurophysiology of emotional language dynamics. Submitted to Neuro-Psychoanalysis.

Crick, F. & Koch, C. (2000a) The unconscious homunculus. Neuro-Psychoanalysis 2(1):3–11.

Crick, F. & Koch, C. (2000b) Response to commentaries. Neuro-Psychoanalysis 2(1):48–59.

de Saussure, F. (1916) Course in general linguistics. Collins Fontana: Glasgow.

Evans, D. (1996) An introductory dictionary of Lacanian psychoanalysis. Routledge: London and New York.

Freud, S. (1953a) On aphasia. A critical study. International Universities Press: New York. Originally published in 1891.

Freud, S. (1953b) The interpretation of dreams. In: Strachey, J. et al. (eds)The standard edition of the complete psychological works of Sigmund Freud, Volume 4 & 5. Hogarth Press: London. Originally published in 1900.

Freud, S. (1953c) Freud's psycho-analytic procedure. In: The standard edition, Volume 7. Hogarth Press: London, pp. 247–254. Originally published in 1904.

Freud, S. (1955) Studies on hysteria. The standard edition, Volume 2. Hogarth Press: London. Originally published between 1893 and 1895.

Freud, S. (1957a) Repression. In: The standard edition, Volume 14. Hogarth Press: London, pp. 139–158. Originally published in 1915.

Freud, S. (1957b) The unconscious. In: The standard edition, Volume 14. Hogarth Press: London, pp. 159–215. Originally published in 1915.

Freud, S. (1961a) The ego and the id. In: The standard edition, Volume 19. Hogarth Press: London, pp. 1–66. Originally published in 1923.

Freud, S. (1961b) A note upon the "mystic writing-pad". In: The standard edition, Volume 19. Hogarth Press: London, pp. 225–232. Originally published in 1925.

Freud, S. (1966a) Project for a scientific psychology. In: The standard edition, Volume 1. Hogarth Press: London, pp. 281–397. Originally published in 1895.

Freud, S. (1966b) Extracts from the Fliess papers. In: The Standard Edition, Volume 1. Hogarth Press: London, pp. 173–280.

Freud, S. (1968) An outline of psycho-analysis. In: The Standard Edition, Volume 23. Hogarth Press: London, pp. 139–207. Originally published in 1938.

Freud, S. (1985) The complete letters of Sigmund Freud to Wilhelm Fliess 1887–1904. The Belknap Press of Harvard University Press: Cambridge Massachusetts and London.

Geerardyn, F. (1997) Freud's project: The roots of psychoanalysis. Rebus Press: London.

Knockaert, V. (2001) Van de luchtspiegelingen van het weten naar de docta ignorantia: Een lezing van "Variantes de la cure-type" [From the mirages of knowing to the docta ignorantia, A reading of "Variantes de la cure-type"]. Psychoanalytische Perspectieven 46:93–110.

Knockaert, V./Geerardyn, F./Van de Vijver, G./Van Bunder, D./Bazan, A. (2002) Anticipation, memory and attention in the early works of Freud. International Journal of Computing Anticipatory Systems. In press.

Lacan, J. (1974) Télévision. Seuil: Paris.

Lacan, J. (1977) The function and field of speech and language in psychoanalysis. In: Lacan, J., Ecrits: A selection. Norton & Company: New York, pp. 30–113.

Natsoulas, T. (1984) Freud and consciousness I: Intrinsic consciousness. Psychoanalysis and Contemporary Thought 7:195–232.

Natsoulas, T. (1985) Freud and consciousness II: Derived consciousness. Psychoanalysis and Contemporary Thought 8:183–220.

Smith, D. L. (2000) Freudian science of consciousness: Then and now. Neuro-Psychoanalysis 2(1):38–45.

Van Bunder, D. (2001) Over het concept structuur bij Merleau-Ponty en in de psychoanalyse [On the notion of structure in the work of Merleau-Ponty and in psychoanalysis]. Submitted to Psychoanalytische Perspectieven.

Van Bunder, D./Knockaert, V./Van de Vijver, G./Geerardyn, F. (submitted) The interaction between psyche and soma: A combined phenomenological and psychoanalytical viewpoint. Paper presented at the 16th World Congress on Psychosomatic Medecine, Göteborg, August 2001.

Van Gulick, R. (2001) Reduction, emergence and other recent options on the mind-body problem: A philosophic overview. Journal of Consciousness Studies 8(9–10):1–34.

Dori LeCroy

An Evolutionary Perspective on the Freudian Concept of Defense Mechanisms

Abstract

By equating the concept of repression with the evolutionary concept of self-deception, the FREUDIAN concept of the defense mechanisms can be considered evolved mechanisms favored by natural selection because they mediate adaptive behavioral strategies. Defense mechanisms may be implicated in the various forms of attachment described by attachment theory and may be evolved strategies that are active during the early years in the context of parent-offspring conflicts and also during adult reproductive years. Examination of the adaptive value of psychoanalytic ideas concerning intra-psychic processes may aid in the exploration of evolved patterns of information organization in the human psyche.

Key words

Self-deception, attachment, parent-offspring conflict, defense-mechanisms, repression.

EVOLUTIONARY psychology posits the existence of complex species-typical psychological mechanisms in humans. With its description of the division of the mind into conscious and unconscious realms and, by virtue of this division, the operation of defense mechanisms, so does psychodynamic theory. I argue that both schools of thought are talking about very much the same thing—that the divisions in the mind and the defense mechanisms in the psychodynamic sense are proximate, motivational causes of evolved, adaptive behaviors, or psychological mechanisms in the evolutionary psychology sense.

Specifically I posit that the FREUDIAN concept of repression and the evolutionary principle of self-deception are the same thing—that like the evolutionary view of self-deception, the human capacity for repression arose because it served individual fitness. In other words repression is the result of natural selection because it was and is adaptive. As David BUSS says: 'All psychological theories, even the most ardently environmental ones, imply the existence of psychological mechanisms', and 'evolution by natural selection is the only known causal process capable of producing complex physiological and psychological mechanisms' (BUSS 1995, p2).

I suggest that the defense mechanisms in the FREUDIAN sense are systems of self-deception in the evolutionary sense, i.e., that they are the result of natural selection, and therefore the defense mechanisms are evolved responses which are underpinned by self-deception (repression); that they are active early in life, functioning to gain nurture from adults distracted by other interests; that they are active life long; and that they are context sensitive and implicated in attachment theory.

Psychodynamic theory refers to the organization of perceptions, thoughts, and feelings within the mind for the purpose of minimizing painful affect, anxiety reduction; in other words, as wholly intra-psychic processes. But to be considered adaptive, these mechanisms must impact the reproduction success of the individual. In a DARWINIAN sense they would refer to overt behavioral tendencies, mediated by emotive and cognitive underpinnings, which were honed by natural selection to meet adaptive challenges during our evolutionary past or in the 'environment of evolutionary adaptiveness' (EEA). Evolutionary psychologists believe that psychological characteristics adaptive in the EEA came to comprise the human mind and we still carry those characteristics with us. Whether they are still adaptive in the modern environment is open to question, but this question is not relevant to the present argument.

We are the descendants of the successful. It was the more fit individuals who survived to become our ancestors while the less fit died out. Fitness is the result of successful negotiation for limited resources: food, protection and mates, and the social position that insures these. At the beginning of life the essential resource is parental indulgence. In the

EEA this was in the face of scarce resources and competing siblings (existing or potential).

For much of our evolutionary past, when our parenting and attachment psychologies evolved, starvation was frequently just around the corner. Average lifetimes were short and infant and child mortality was high. A woman spent most of her time pregnant or lactating. Commonly she was toting a toddler along with her infant while gathering the foods that supported them all. All this makes childbearing energetically very expensive.

Parent–offspring conflict theory states that with only 50% relatedness the interests of parents and children imperfectly overlap (TRIVERS 1974). In this genetic economy parental interest is more in lifetime reproductive success than in any individual offspring. Thus parental interests conflict with that of particular offspring, sometimes, and frequently during our evolutionary past, with fatal results. Contemporary hunter-gatherers still allow ill-timed, unhealthy or "wrong sexed" newborns to die of neglect (HRDY 1999). So from the start, the individual human is challenged to survive. Throughout history, exposure, unhealthy wet nursing conditions, foster parenting, child labor, over-crowded orphanages have been prevalent.

Unlike other primates, human mothers are not unconditionally committed to every newborn. Sarah HRDY (2002) believes there is a 72 hour period, the post birth-pre-lactation period, of maternal ambivalence during which a decision is made. Environmental conditions, resource availability including social support and the health of the neonate are what tip the balance. HRDY further suggests that human newborns, unlike other primates, are extra chubby. The addition of fat onto the baby during the last few weeks of pregnancy is calorically expensive to the mother. Its uniqueness among the primates makes it seem a gratuitous expense from the maternal point of view so HRDY speculates it is an offspring strategy in that the healthy appearance of a fat baby makes it seem worth further investment (HRDY 1999). In addition, during childhood when infanticide is no longer an issue, neglect because of dire circumstances or the existence of a favored sibling might be, and was a likely occurrence in the EEA. Even today there is evidence that parents favor offspring with the greatest potential for survival and eventual reproduction (BUSS 1999).

Thus it appears that, in the service of a woman's fitness, there are evolved mechanisms activated by some kind of primitive unconscious assessment of unpromising reproductive circumstances which are capable of overriding the mechanisms of maternal devotion. If this is true, then by the same rules, we are bound to consider natural selection's role in equipping offspring with their own strategies to counter sub-optimal parenting. Baby fat may be one of these. We also have the infant's seductive gaze and endearing smile. But what of others, and most interestingly, what psychological ones might exist in the child at the mercy of parents who must make decisions about the distribution of limited resources? Surely natural selection would not have left them passively awaiting their fate.

A look at attachment theory with an evolutionary eye suggests that the varieties of attachment styles, secure and insecure, reflect the influence of natural selection. The term secure attachment has been used to designate a mother–child bond marked by consistently responsive parenting style and a child able to explore his environment confident of parental protection and solicitude (AINSWORTH et al. 1978; BOWLBY 1973).

Using the AINSWORTH Strange Situation Test (AINSWORTH et al. 1978), attachment theorists identify varieties of infant behaviors as secure and as two varieties of insecure, avoidant and resistant. These are behavior patterns that are systematically traceable to the quality of early parental care. This categorization of infants is based on their reaction to separation from their mother. Secure infants may or may not be mildly distressed upon separation, and they show pleasure at reuniting with the parent. They have parents who consistently respond to their needs. Infants with insecurity of the avoidant type show neither distress when separated from, or pleasure when reunited with, the mother. Their parents are consistently unresponsive to them. Insecure/resistant types of children are very upset upon separation and generally cling to their mothers. These children have parents who are unpredictable.

It has been found that the style of attachment during early life is correlated with that of later life, including the way it manifests in male–female pairbonds, e.g., short-term and exploitative, or enduring and mutualistic (GOLDBERG 1997). Recently the idea that all the variants of attachment style may have been selected for because they are adaptive in different circumstances is being explored. Cross-cultural studies indicate that the prevailing nature of male–female attachment in a culture is correlated with the reliability of resources. To some this suggests that attachment styles of all kinds constitute a repertoire instantiated by natural selection to guide the behavior of the reproducing adult. Jay BELSKY states that, "…the attachment system evolved as an environmentally contingent mechanism for promoting reproductive fitness in adulthood…". He suggests that the varieties of insecurity may be adaptively specific to local conditions (BELSKY 1997). So BELSKY sees the youthful versions of attachment style as preparation for successful reproductive strategies in later life. What is less explored is the fact that the varieties of attachment style may also be specific offspring strategies that benefit the individual while still very young.

One might fruitfully consider the varieties of attachment style as coping strategies during childhood, at least under conditions like that of the EEA, in addition to their adaptive value during later reproductive life. Natural selection may have installed a range of attachment styles in the very young that are variously expressed according to immediate conditions—particularly conditions similar to ones prevalent in the EEA.

The evolutionary principle of parent–offspring conflict suggests that the varieties of attachment styles make sense as offspring strategies for attaining resources. Avoidant/insecure attachment represents low expectations of nurture from parents and reduction of the wasted effort of continuing to make appeals in that direction. (This may seem implausible for contemporary children living in the family isolated patterns we are used to seeing. However, in the EEA, when our evolved behaviors were shaped, it is likely that grandmothers and other kin lived in a common space within hearing and seeing distances.) Alternatively, the resistant/insecure type may be an adaptive parent-centered strategy based on behaviors that reduced offspring cost to parents. We will return to this pattern in more detail below. The immediate adaptive consequences of the various attachment styles and an evolutionary view of attachment systems in general is now beginning to be considered by attachment theorists (e.g., SIMPSON 1999).

Attachment theorists have designated the terms 'secure' and 'insecure' to identify the classes of behavior they observe. Note that these are attitudinal terms not behavioral ones. This presumes the existence of influences that set up attitudinal biases, during early life. Such biases would be based on internal models of the world that emerged from the consistency and appropriateness of responses by parents to a baby's needs and desires. In other words, patterns of expectations are set up by early experience of nurture and they influence the selection of tactics for the manipulation of others and for the manipulation of the self (repression, self-deception) in ways that foster the acquisition of resources.

The terms, 'secure' and 'insecure' would then represent intra-psychic processes that reflect the action of defense mechanisms as described by FREUD. It is these defense mechanisms that are expressed in the attachment styles within the parent–offspring relationship during early life and remain active in the male–female relationships later on.

FREUD's idea of the function of psychodynamic processes was mediation between inborn drives and social expectations. It is an intra-psychic conflict model, but a little retooling can fit it into the theory of parent–offspring conflict. If maternal reproductive strategies in the face of limited resources can include decisions that can result in the neglect of offspring, then natural selection would have encouraged offspring behavior to cope with this threat.

For example, displays of distress and anger are the usual ploys of frustrated children. They work to gain resources from parents. However when resources are scarce and parents are ambivalent demanding tactics may prove fruitless, even hazardous, as they could provoke parental aggression. But complete elimination of offspring appeals could lead to serious deprivation and different ploys would be useful. Thus repression, in the psychoanalytic sense, or self-deception in the neo-DARWINIAN sense, may have been natural selection's "solution".

Repression of the perception of parental withholding and emotional reactions to it would permit strategically advantageous behavior impossible otherwise. In other words, psychological disconnection from grief and rage, or repression/self-deception, permits the deployment of other more successful behaviors like those that influence nonparents to provide resources or behaviors that can better seduce ambivalent parents into increased nurture.

Instead of showing anger and resentment, the children of inconsistently responsive parents may become solicitous, overly helpful and doting. This leads to situations familiar to clinicians where children are parenting parents. They take on parental tasks like household duties, preparing food, caring for siblings. These behaviors are common in hunter-gatherer societies and presumably in the EEA as well. They are all tactics that reduce offspring cost to parents. Through these behaviors the acquisition of resources, like food and protection, not otherwise forthcoming because of its scarcity or of parental favoring of a sibling, may be attained.

Doting service to parents who are withholding has to be the result of a way of thinking and feeling. Motivation by a psychic organization of some sort has to be assumed. I see psychoanalytic concepts as likely mechanisms for this psychic organization. For then we can talk about splitting, idealization, denial, projection, identification and reaction formation, all the defense mechanisms, as the mechanics of the self-deception that must be operating. There is, after all, a huge body of contemporary psychoanalytic thought that has moved beyond its classical origins through the careful observation of human behaviors.

In my clinical work, I saw several individuals who idealized and patronized withholding or abusive parents. Indeed I can think of few who did not idealize one or both parents. From the psychoanalytic viewpoint we can see this as self-deception in the service of splitting, identification and idealization. All of which are considered normal processes, but exaggerated in certain individuals such as might be the case with insecure attachment styles. The scenario goes like this. The maternal figure that is sometimes satisfying and sometimes frustrating is perceptually processed by splitting off the frustrating aspects. This achieves a cleansing of the maternal figure into a wholly good, ideal mother deserving of adulation. The frustrating figure is then internalized as personal 'badness'. The psychoanalytic explanation is that this creates a more comfortable, hopeful view of reality than would be the case if one must believe in maternal treachery. This has been called the 'moral defense' (FAIRBAIRN 1952) and it may bring psychic comfort, but my point is that it may also motivate adaptive behavior.

Psychoanalytic theory postulates that all children split the maternal figure and internalize guilt and shame rather than see parents as unsatisfactory. Adults do the same thing. It is well known that battered women feel they deserve mistreatment and suffer from "low self-esteem". They do not see the perpetrator as blameworthy but rather they defend him instead. They justify the mistreatment they suffer by blaming themselves. They reason that the abuser is forced to punish them because of their own misconduct and that this is a reflection of the goodness of the perpetrator whom they idealize. In other words, self-deception orders thoughts and feelings in situations where individuals are dependent on dominants that are withholding or abusive.

It seems clear we have not evolved a cognitive system all in one piece and designed for an unbiased perception of reality. Rather, incoming information is processed in ways that variously represents reality. It also seems that distortions of reality are systematic and I argue that these are evolved and based on self-deception.

To quote Robert TRIVERS (2000):

"An important component of a mature system of social theory is a sub-theory concerning self-deception (lying to oneself, or biased information flow within an individual, analogous to deception between individuals)… whether through study of one's own behavior and mentation…or a review of findings from psychology, we know that processes of self-deception—active misrepresentation of reality to the conscious mind—are everyday occurrences, that struggling with one's own tendencies toward self-deception is usually a life-long enterprise. …A theory of self-deception based on evolutionary biology requires that we explain how forces of natural selection working on individuals and the genes within them may have forwarded (it)."

The question TRIVERS suggests here is: How could self-deceiving individuals thrive better than those whose entire minds were devoted to accurate, faithful renditions of reality? Or how is it that self-deception mechanisms are adaptive? His answer is not that it brings psychic comfort but that advantage is found in deceit. In other words, the adaptive advantage of self-deception is improvement in success in deceiving others.

For example, your supervisor has decided to promote a colleague instead of you. The reason he gives is that you will have time for special training that will be of greater ultimate advantage for you. In spite of evidence that he is really acting out of self-interest you might benefit from consciously believing his story. Self-deception in this case would allow you to uphold an authentic appearing, good-humored deference, and you would avoid the costs that might be incurred if suspicion and resentment should reveal themselves to your superior. In addition, for his part, your boss also misleads himself by really believing his decision is in your interest. Thus allowing him to behave convincingly toward you and also get a bonus of feeling beneficent.

Also consider how much more effective a sexual seduction tactic is when the male believes himself utterly smitten. The cold light of the morning after might bring a more realistic view, but by that time the seed is sown. A seed, not incidentally, that contains the genetic contribution to effective self-deception.

Now consider this. A twelve-year-old devotes herself to anticipating and responding to her mother's needs and desires. All in spite of the fact that her mother is frequently cold and withholding. Here is self-deception in support of a reaction formation where over-solicitous behavior is overlying a disappointment and resentment that remains beneath consciousness. But her behavior is convincing because she believes it, and mother indulges her as a result.

The focus of natural selection would have been on this outcome, maternal indulgence, as it would be on indulgence by non-parents in the case of the insecure/avoidant type who has lowered expectations concerning parents and looks to others. But what would the internal organizing mechanisms have been? Behavior emerges out of habits of perception, thought and feeling, conscious and unconscious, yet psychoanalytic and allied schools of thought that are specifically concerned with just such intra-psychic phenomena are largely overlooked for motivational concepts worthy of consideration within an evolutionary framework.

The forgoing is an example of how psychoanalytic constructs like the defense mechanisms can be reinterpreted as motivators of adaptive behavioral strategies. This approach unifies aspects of attachment theory with the evolutionary principles of parent–offspring conflict and the theory of self-deception, and brings these into relation with psychoanalytic concepts. In a similar way, we might look at the adaptive impacts that might be consequent to the other defense mechanisms. For example let us briefly look at the defense mechanisms of projection, identification, and regression.

The classic example of projection is the case of the individual whose public contempt for homosexuality arises out of reaction to his own unconsciously perceived, and disapproved, homosexual feelings. By his projection of homosexuality outside himself he is self-deceiving and avoiding painful anxiety and shame. But given that the neo-DARWINIAN view of self-deception is that it is in the service of deceiving others, this instance of projection might be seen as a socially beneficial advertisement of "virtue" through public proclamations of another's guilt, which both diverts attention from one's own shame and also enhances social status. In human groups status is key to resource acquisition, which is the same as saying the key to fitness, at least in the EEA.

Identification refers to processes by means of which an individual integrates aspects of another's personality. Through identification with an admired or powerful individual, one may incorporate adaptive strategies. One type of identification described by psychoanalytic theory is 'identification with the aggressor'. This occurs when a victim of mistreatment mitigates the intensity of their distress by identifying with the actions of the mistreating individual. From a neo-DARWINIAN perspective, identification with the aggressor may be seen as a mechanism that encourages submission to others when it is strategically advantageous to be a follower rather than waste energy on futile clamoring for dominance. Also, association with dominant others brings benefits not available when one remains unassociated.

Regression is a process by means of which one assumes a stance of reduced autonomy and competence. In the psychoanalytic view it is a form of retreating to a less anxiety evoking level of functioning. However, it is a social as well a personal act, and may be effective in soliciting goods and services when fitness needs are frustrated as may be the case when one feels unsupported, exploited or dominated. Thus, energy conservation and some social welfare, rather than psychic relief may be the advantage that selected for regression.

FREUD emphasized psychic relief as the function of defense mechanisms rather than social advantage. From an evolutionary perspective, varieties of psychic discomfort are internal signals indicating that circumstances are not serving one's strategic goals or fitness and also that they are motivators of adaptive responses. With an evolutionary perspective we may find that the storage and organization of information outside of awareness, or in what FREUD called 'the unconscious', and the resultant defense mechanisms, constitute evolved mechanisms that foster adaptive behavior. In other words, we may find that examination of the defense mechanisms as described by FREUD would be a good starting point for the exploration of evolved patterns of information organization in the human psyche.

Author's address

Dori LeCroy, 503 N. Broadway, Upper Nyack, New York 10960, USA.
Email: dorilecroy@aol.com

References

Ainsworth, M. D./Blehar, M. C./Waters, E./Wall, S. (1978) Patterns of attachment: A psychological study of the strange situation. Erlbaum: Hillsdale NJ.

Belsky, J. (1997) Attachment, mating and parenting: An evolutionary interpretation. Human Nature 8(4):361–381.

Bowlby, J. (1973) Attachment and loss: Vol. 2. Separation: Anxiety and anger. Basic Books: New York.

Buss, D. M. (1995) Evolutionary psychology: A new paradigm for psychological science. Psychological Inquiry 6(1):1–30.

Buss, D. M. (1999) Evolutionary psychology: The new science of the mind. Allyn & Bacon: Boston.

Fairbairn, R. (1952) Psychoanalytic studies of the personality. Routledge: New York.

Goldberg, S. (1997) Attachment and childhood behavior problems in at risk and clinical samples. In: Atkinson, L./Zucker, D. J. (eds) Attachment and psychopathology. Guilford: New York, pp. 171–195.

Hrdy, S. (1999) Mother nature: History of mothers, infants, and natural selection. Pantheon: New York

Hrdy, S. (2002) Plenary Address at the XVIth Biennial Conference of the International Society for Human Ethology, Montreal, Canada.

Simpson, J. (1999) Attachment theory in modern evolutionary perspective. In: Cassidy, J./Shaver, P. (eds) Handbook of attachment: Theory, research, and clinical applications. Guilford: New York, pp. 115–140.

Trivers, R (1974) Parent–offspring conflict. American Zoologist 14:249–264.

Trivers, R. (2000) The elements of a scientific theory of self-deception. In: LeCroy, D./Moller, P. (eds) Evolutionary perspectives on human reproductive behavior. Annals of the New York Academy of Sciences 907. New York Academy of Sciences: New York, pp. 114–131.

Gertrudis Van de Vijver/David Van Bunder/Veroniek Knockaert/
Ariane Bazan/Filip Geerardyn

The Role of Closure in a Dynamic Structuralist Viewpoint of Psychic Systems

Abstract

In this paper, the emergence of psychic structures is conceived of within a dynamic structuralist framework, which is developed in dialogue with the biological viewpoint on the emergence and the maintenance of living structures (COLLIER/HOOKER 1999; SALTHE 1985; VAN DE VIJVER/SALTHE/DELPOS 1998). More in particular, it is suggested (i) to consider organizational closure as a minimal condition to be fulfilled in order for dynamic systems to be stable, self-maintaining and self-producing, (ii) to conceive of living systems in terms of intricate organizational closures that set the stage for meaningful interactions (KAUFFMAN 1993; MATURANA/VARELA 1980; PATTEE 1995; ROSEN 1985, 1991; VARELA 1979) and (iii) to situate the psychological realm in continuity with this viewpoint.

In analogy with the organizational closure of living systems, the mechanism of psychic closure is presented as the key mechanism for the constitution of psychic structures, and is interpreted in terms of identification: an identificatory judgment is a self-referential, closing judgment that involves a reinterpretation of the level below (the organic body) and is addressed to the level above (the fellow human being). The way in which identification has worked determines the kinds of psychic structures that are formed, it leads to different kinds of signifying practices, as well as to different ways of taking into account the underlying biological and material processes.

Key words

Closure, identification, complex dynamical systems,

Introduction

Few people would deny that human beings are capable of entertaining meaningful relations with their environment. Humans are thought to be among the most flexible in this regard, more flexible than other animals, and in any case radically different from non-living systems. This is almost common knowledge, but do we know exactly what these human signifying capacities involve? Do we know at what levels meaningful interactions can take place? Do we know how to distinguish precisely between conscious and unconscious interaction? Do we know how signifying capacities arose in a developmental and evolutionary history? Do we know the basic differences on these points, if any, between human and other living systems?

There have been many attempts to answer these questions, but the answers have varied greatly, depending, amongst other things, upon the interpretation of what is meaning, on the definition of a system or a living system, or on the developmental and evolutionary antecedents considered as relevant. As I shall be dealing to a great extent with the problem of meaning, and in particular with the structural and dynamic conditions within which meaningful interactions can emerge, in biological as well as in psychic systems, let me begin by giving a rough sketch of the background assumptions that determine my questions.

Firstly, there is an intricate relationship between what makes living beings living, and what makes them capable of acting and behaving in meaningful ways (cf. HOFFMEYER 1996; VAN DE VIJVER 1997). If we wish to understand signifying processes, we need to understand signifying processes in living systems. This is the core-idea of *biosemiotics*, that focuses on the informational, communicative and semiotic aspects of living entities and processes (cf. HOFFMEYER/EMMECHE 1999), and proposes to transgress the borders of human sign-practices and human-centered ways of understanding, by considering them as part of signifying practices of living systems at large.

Secondly, if living beings are capable of signifying practices, it is because they are *complexly organized* *dynamical systems*. In other words, processes of meaning are to be related to the constraining influence of certain dynamical structures. This viewpoint comes from *the theory of complexly organized dynamical systems* (COLLIER/HOOKER 1999; CHRISTENSEN/HOOKER 2000), also called *dynamic structuralism* (VAN DE VIJVER 2000) *or interactivism (*BICKHARD 2000). This approach considers that structures emerge out of dynamic processes, and focuses on the levels, the moments and the modalities of this process of emergence, as well as on the constraining impact these structures can have in return on the dynamical processes. It considers living systems as the more or less fixed, constrained and constraining, emergent results of particular dynamic histories. Living systems have a global, constraining impact because they are cohesive wholes that globally determine the scope and the meaningfulness of the local interactions. In other words, the actively maintained cohesion of living systems has a determining, constraining, selective, interpretive impact on the environment, which no longer dictates in itself the ways of being interpreted. This is, in more philosophical terms, what intentionality is about. Intentional behavior—the intentional directedness towards something—is constructed in a developmental context as the active, interpretive impact of global structures. It is embedded in multi-layered systems, not restricted to consciousness, and is grounded in evolutionary intentionality, i.e., the anticipative power inherent in all living systems (HOFFMEYER 1996, p47).

Thirdly, the *emergence of psychic structures*, characteristic of human beings, is to be conceived of in line with the biological viewpoint on the emergence and maintenance of living structures (VAN DE VIJVER 1999, 2000). The psychic system is seen as a complexly structured dynamical system with a particular evolutionary and developmental history. As such, it determines the scope and the meaningfulness of the interactions with its environment, including the environment constituted by the own body. Various psychic structures can emerge depending upon the structural and developmental conditions. Here, a connection can be made with FREUDIAN and LACANIAN psychoanalytic theories, that have a very similar account of psychic structures. Even if the evolutionary embeddedness is rarely acknowledged in this context, the developmental history and the constraining impact of psychic structures is clearly of major importance for psychoanalytic theory and praxis.

And finally, an adequate account of meaning can only be *a relational or triadic* one, in the sense of Peirce's pragmatism. In stating that "a sign (…) is something which stands *for* something *to* somebody in some respect or capacity" (PEIRCE 1897, in BUCHLER 1955, p125), PEIRCE clearly stressed that if there are "signs of meaning" in the universe (HOFFMEYER 1996), they are always signs *to* someone, never signs as such. Only when there is a "someone" capable of interpreting something as meaningful, can there be signs. As a consequence, the Peircean sign-conception is committed to a universe populated by dynamical, active, "subjective" systems, i.e., beings that have developed some kind of autonomy or agency out of which they are able to take something as standing *for* something else. Hence, there are no signs as such, and no meanings as such. Meaning is realized *in between*, in the dialectics between the sign vehicle (including representations or symbols or language), the interpretant or processing activity (wherever it is located), and the object or the referential aspect (the aboutness). Meaning is neither in the representations, nor in the things themselves, nor in the processing activity as such. Meaning is the *"in-between"* itself.[1]

With these minimal clarifications, let me take up again my initial questions and focus them a bit more. Understanding what the capacity of entertaining meaningful relations involves in humans, requires an understanding of: (i) what conditions have to be fulfilled for systems to be complexly organized and dynamical, (ii) how these are linked to the possibility of signifying practices, (iii) how these conditions are fulfilled in psychic systems.

What I want to focus on in this paper, is the issue of closure, its role in living and in psychic systems for the constitution of dynamical structures, and its implications for the issue of meaning. It is suggested (i) to consider *cohesion*, a specific form of closure, as a minimal condition to be fulfilled in order for dynamic systems to be stable, self-maintaining and self-producing (COLLIER, in press), (ii) to conceive of living systems in terms of *intricate organizational closures* that set the stage for meaningful interactions, and (iii) to situate the psychological realm in continuity with this viewpoint. If closure, and in particular cohesion, is considered as the central mechanism for the formation and maintenance of living structures, then our task will be to understand the status, the meaning and the implications of closure at the psychological level. What can be meant by psychic closure? What makes psychic systems into systems? What mechanisms lead biological systems to close themselves organizationally and to become psychic systems of a certain kind? Are there various kinds of psychic systems, differently structured? How do psychic systems acquire the possibility of entering into meaningful interactions? What are the consequences for our understanding of the unconscious?

The Role of Closure in Complexly Organized Dynamical Systems

1.1. A Classification

Within the frame of complexly organized dynamical systems, the following classification of systems is proposed:[2]

	Non-organized	Organized
Simple	Type 1 *single particle, conservative, decomposable (linearisable) multi-particle systems*	Type 3: Static organization, sufficiently well constrained but non-linearisable multi-particle systems *some machines, some electro-magnetic systems*
Complex	Type 2: Statistically specified at or near equilibrium *gases, fluids*	Type 4: Dynamic organization *hurricanes, cells, living, autonomous systems*

Complexly organized systems, unlike systems of type 1 and 3, have an internal organization which is not fully determined by initial internal and boundary conditions, but which can change through time. Living systems are said to be complexly organized dynamical systems: their internal organization is not totally imposed by boundary conditions (such as in complex but unorganised systems), neither is it fully determined by initial internal and boundary conditions (such as in unorganised and simple systems). The internal organization of living systems is, on the contrary, the result of an active and continuous process of self-maintenance, self-organization and self-reorganization. Self-maintenance and autonomy are thus processes that actively and continuously construct and maintain a difference between the system and its surroundings, between "self" and "non-self". This creation and maintenance of a difference presupposes an integration over various processes at various levels, and expresses a dynamical and interactive process in which living systems are not passively determined by the environment, but actively interpret and co-construct the interactions with their surroundings. So, the behavior of complexly organized dynamical systems is neither a linear function of the inputs they receive, nor of their internal structure.[3]

One of the most intriguing properties of living systems is their *integrated, global, cohesive, non-decomposable* character, which makes that we describe them as wholes that cannot be significantly reduced to the workings of the parts. As Collier and Hooker state: "(…) autonomous systems are intrinsically organizationally global (…) Because their capacities, i.e., their processes and resulting functional properties, must be so *integrated* that they are able to actively maintain themselves, their overall functionality cannot be grounded in a mere aggregate of independent processes but requires that distinctive global *process integration* that alone insures maintenance of the whole as a joint interactive consequence across all their interrelated process cycles. Process control in such integrated systems is typically complex, acting across many different timescales (…) and requiring *coherent* activation and modulation (…) of subsidiary processes." (Collier/Hooker 1999, p6, italics added).

Two questions will be addressed here: (i) what are the specific structural conditions to be fulfilled to obtain such global, integrated wholes, and (ii) on what grounds do we say that something is integrated and global—is it an intrinsic property of complex systems or are we implied as observers?

1.2. Circular causality, organizational closure, cohesion

1.2.1. Structural conditions. To account for the integration and autonomy of living systems, many philosophers as well as scientists have introduced the idea of a *circular causality*, i.e., of a closure of the causal relations between the components of a system.[4] Among the most noticeable are Maturana and Varela, who introduced the concept of organizational closure, which refers to the self-referential organization of the causal interactions in material systems, i.e., to the self-referential, recursive organization of the causal loops that determine their dynamics.[5] In this view, organizational closure clearly does not imply that the system under consideration is closed to energetic or material interactions with the environment. An organizationally closed system is a network of dynamical processes of which the effects remain within the network. The main merit of this approach is to have highlighted the capacities of *self-production* and *self-maintenance* of living systems, and to have linked these explicitly, in cybernetic terms, with the self-referential or circular organization of the causal relations between the components. However, it has the drawback to have focused entirely on the product or the result of processes of closure, to stress mainly the "self-contained" or solipsistic character of living systems, and to neglect their multi-layered character.

The theory of complexly organized dynamical systems provides us today with a more encompassing and more adequate account of living systems, in at least two ways. Firstly, this theory expands the idea of organizational closure to multi-layered and multi-interactive systems and introduces to that end the concept of cohesion, which is a particular form of organizational closure. Secondly, it avoids problems of solipsism by articulating the idea of interaction at various levels, within the system and between the system and the environment, and by linking the possibility of meaningful interaction to the cohesive nature of complex systems. As we will see, this will have some resonance in the domain of psychological phenomena.

The main concept for our purposes is the one of cohesion, which is a form of closure "(…) of the causal relations among the dynamical parts of a dynamical particular that determine its resistance to external and internal fluctuations that might disrupt its integrity. [It includes] the requirement that these relations be stronger on average within the closure than without. This determines a *cohesion profile* that gives the (probabilistic) conditions under which a thing will both retain and lose its integrity, determining its boundaries under a range of conditions. We thus describe cohesion as the "dividing glue" of dynamical entities. (COLLIER, in press, p4).[6]

As in any process of organizational closure, the closure is dynamical or causal, realized in physical processes: "cohesive constraints are dynamically real, they arise from actual dynamical interactions of the system components, and once formed they genuinely constrain the behaviors of those components" (COLLIER/HOOKER 1999, p17). Moreover, cohesion is active, flexible and holistic. In the case of living, autonomous systems, cohesion arises out of the interactions with the environment which are part of a causally integrated totality. As can be seen, cohesion involves *process- and interaction closure*: cohesion rests on an interactive process that depends on system- as well as environmental factors. Process closure refers to the fact that the processes need to bring forth the conditions under which the system perpetuates. In order to attain global process closure, the elements of the system have to interact with the environment and with each other in specific, self-sustaining ways. Interaction closure refers to the specific interactions the elements entertain with the environment and that are causally closed in the sense that they promote the maintenance of the conditions that generate these interactions.

A consequence of the global, constraining character of cohesive systems, is that they have the capacity to *differentiate*, to switch, between different environmental conditions and hence, between different types of interactions.[7] To a certain extent, they can *choose* the interactions they consider as relevant or meaningful. Here, an intimate connection is to be made with the issue of semiosis: cohesion provides us with the relevant context that co-determines the meaning and the scope of the interactions at play in the system. Without cohesion, no context of interpretation. If living systems are able to interpret their environment, it is, minimally, because they are organizationally *closed, cohesive* systems. Cohesion provides them with a form of stability, protecting them from being invaded or destroyed by stimuli of all kinds, and leaves openness for new, potentially meaningful stimuli that arise from contacts with the world. This relative stability, expressing the capacity to actively resist internal and external fluctuations, together with this relative flexibility, expressing the capacity to selectively determine and anticipate the impact of the stimuli, constitutes the core of the interpretive context of living systems.

1.2.2. Epistemology. This brings me to the second question: on what grounds do we say that something is cohesive—is it an intrinsic property or are we inherently implied as observers? Interactivists are keen to suggest that closure refers to the *causal* relations between the material parts of the system, thereby suggesting a total absence of functional concepts. However, their viewpoint is not totally convincing, for at least two reasons.

Firstly, in their view, the knowledge of causality itself is always inferential and abductive—you cannot observe causes, you have to infer them. As a consequence, the process of inferring causal circular processes is intimately linked to the use of *functional* "self"-terminology, such as "self-sustaining", "self-organizational", "self-maintaining". Is there a means of dealing with circular causality in completely non-functional terms? In my opinion, the answer is no.

Secondly, what about the assessment of any type of cohesion or closure? Is it not the case that this assessment inevitably rests upon the perception of so-called autonomous systems; is not that assessment determined by the interactions we can have with them, interactions in which self-movement, relative stability and relative flexibility will play a prominent role? Is that assessment not constrained by the interests and purposes, by the choices and wishes, characteristic of the interactive systems that we are? Kant was very clear on this point: in assessing living systems, we have to consider them as purposive wholes, this is the meaning we need to add, otherwise we would be completely unable to make sense of them. Their wholeness captures their essence, but this is and remains for ever a non-objective property, related to our ways of interrelating, interacting and communicating with living systems. It is possible to neglect their character of purposive wholes, but then, can we still understand them?

Let us now see in what way we can connect these issues with the emergence of psychic structures. As will become clear, expanding the theory of complexly organised dynamical systems to the psychological realm is a tricky task, because it complicates the issues and colours them so differently that one sometimes wonders what the use is of putting them in parallel. The problem here is related to what is called 'emergence'. We have been dealing until now with the emergence of structures out of dynamical processes, considering living structures as the more or less fixed, constrained and constraining, result of such an emergent process. Dealing with psychic phenomena or psychic structures, in dialogue, even in continuity, with the biological order, suggests that those structures equally emerge out of underlying dynamical, biological processes. It even might suggest that organizational layers (chemical, biological psychological, …) are in a sense fixed, that their accessibility remains unchanged no matter from what perspective they are studied. This is, however, not the case. Indeed, what the study of psychic structures illustrates most clearly, is that the "addition" of a psychic level itself implies profound shifts in the accessibility of the surrounding levels out of which it emerged. In other words, the issue is not just to arrive at grasping the formation of psychic structures out of, let us say, "underlying" (biological, organic) structures and levels. The issue is to situate the emergence of psychic structures within a dynamic context, that is, to understand how psychic structures are constrained by surrounding, lower and higher, levels, but also to understand how they constrain themselves the interpretation of these levels. In a sense, we need to understand how the psychic level *reinterprets* the surrounding levels, that is, how the constitution of new types of structures on the basis of organizational closure is, as such, a fundamentally semiotic operation.

In what follows, I shall focus again on the concept of closure, leaving many other epistemological issues with regard to emergence untouched (see VAN DE VIJVER 1998).

Psychic Structures: In Parallel with Biological Structures

If we adopt the interactivist frame of thinking, a psychic structure refers to the particular psychic organization that is formed in a particular history on the basis of a particular kind of closure of the causal relations among certain dynamic parts of a material system. What kind of closure can be at work here? What kind of relation can there be between a psychic and biological system? As we have seen, it is necessary to define the mechanisms at work in the construction of psychic structures, and to make explicit the ways in which these structures are determinative for particular signifying practices. The basic idea is that if psychic systems, *qua* living systems, are able to interpret their surroundings, it is because they are a specific type of organizationally *closed*, that is, *cohesive* systems. So, what kind of mechanism is likely to be at work in psychic closure?

In analogy with the cohesive closure of living systems, the mechanism of psychic closure is presented as the key mechanism for the constitution of psychic structures, and is interpreted in terms of identification: an identificatory judgment is a self-referential, closing judgment that involves a reinterpretation of the level below (the organic body) and is addressed to the level above (the fellow human being). The way in which identification has worked determines the kinds of psychic structures that are formed, it leads to different kinds of signifying practices, as well as to different ways of taking into account the underlying biological and material processes.

Identification concerns the way in which a human being succeeds in recognizing itself, in building a representation of itself, or in giving itself a place in the outer world, on the basis of an interaction with its surroundings (GRUNBERGER/CHASSEGUET-SMIRGEL 1978; BOUDAILLEZ 1998; LACAN 1961–62). The basic idea behind identification is that a human being does not stand in a relation of cognitive self-evidence, but needs to develop a sense of self by orienting itself in the world, that is, by actively making a distinction between inside and outside, self and non-self. Topological barriers or membranes (i.e., organic barriers) are not sufficient, even to some extent irrelevant, to provide such a distinction: they indicate it only from an external point of view, and do not necessarily correspond with the point of view that the perceiving system needs to build for itself. For human beings, the development of this distinction takes place progressively, and goes hand in hand with an emancipation from the purely biological order. This is what the process of emergence is about: structures emancipate themselves from other structures, they acquire some sort of autonomy on the basis of progressive differentiation.

Within psychoanalysis, there have been various ways of describing this process of differentiation. FREUD's metapsychology shifted over unconscious, preconscious, conscious, to the theory of the Ego, the Id and the Super-Ego (see VAN BUNDER et al. 2002). LACAN, who disliked developmental terminology, nevertheless frequently suggests, beyond a logical perspective, a developmental one, where he distinguishes between the Real (R), the Imaginary (I), the Symbolic (S). In our treatment of identification, we shall basically adopt LACAN's RSI distinction.

In order to understand the idea of differentiation, let us remain for a while with the question of identification in a being which we assume is as yet totally non-oriented in the world, non psychically structured being, so to speak. What does it mean for such a being to identify or realize itself as a self, something different, something cohesive, something that "stays together as a thing", as FREUD states in the Project? It means that it needs to find an orientation in the world, distinguishing an inside from an outside. A crucial question is then: what will be the element or the process that is likely to reveal to it its nature, its place in the world; or in other words, what will the living being recognize itself in? How shall it know that there are things that belong to "itself" and things that are outside of the scope of its being?

FREUD had in this regard a quite interesting proposition that, although it sketches a mythical starting point, is useful for our purposes. In "Instincts and their vicissitudes", he states the following: "Let us imagine ourselves in the situation of an almost entirely helpless living organism, as yet unoriented in the world, which is receiving stimuli in its nervous substance. This organism will very soon be in a position to make a first distinction and a first orientation. On the one hand, it will be aware of stimuli that can be avoided by muscular action (flight): these it ascribes to an external world. On the other hand, it will also be aware of stimuli against which such action is of no avail and whose character of constant pressure persists in spite of it; these stimuli are the signs of an internal world, the evidence of instinctual needs. The perceptual substance of the living organism will thus have found in the efficacy of its muscular activity a basis for distinguishing between an 'outside' and an 'inside'" (FREUD 1957, p119).

What can we learn from this passage?

The general message seems to be: that which I cannot escape from, what FREUD calls the constant stimulus, that is me, and that which I can escape from, that is not me. Hence, the system identifies with the constant stimulus, or, so it could be stated, it recognizes itself in the stimulus, and on this basis "realizes" its integrity or its cohesion by distinguishing between an inside and an outside (BALDINI 1990; DALTO 1998).

Moreover, FREUD comes to a conclusion that is in various respects remarkable: whereas identification concerns the way in which the "perceptual substance of the living organism" acquires a sense of self, it appears that one always identifies with something that is a priori foreign. One always identifies by "internalizing" an element that is taken as belonging to a self that is thereby constituted. The primitive, initially non-oriented being is not the constant stimulus, it takes the constant stimulus as an indication for the fact that it belongs to itself. As we have seen, FREUD is very clear on this point: "all that I can directly withdraw from, this is not me; on the contrary, all that I cannot directly withdraw from, that is me". It is not the inverse: it is external, and therefore I can withdraw from it, or it is internal, and therefore I cannot withdraw from it.

Of course, the idea of a totally non-oriented being is a mythical one, serving only theoretical purposes in FREUD's reasoning. Biological beings are already not unoriented at birth, even if it is true that they quickly acquire a further orientation on the basis of their movements and the feedback they receive from moving around. This is the basis of what could be called "Real" identification, in the sense of LACAN, a process in which the proper body acquires a more cohesive character, that is progressively identified as something that "moves together", on the basis of interactive feedback processes, in which encounters with so-called "external', foreign elements is crucial. What is identified through this interactive process, is the perceiving and moving body itself. Real identification in this refers to the process- and interactive closure grounded in the movements and perceptions of the body. The acquired cohesion witnesses of a differentiation between the own movement and the other movements, the own perceptions and the other perceptions. Already this type of differentiation is realized through the semiotic operation of taking an element as standing for something else, and shows that this operation rests on the initiated movement from within, the confrontation with an external element, and the feeding back of the effects of this confrontation to the source of the movement.

This type of "nearby" feedback, in which the interpretation of the stimulus quite directly returns to the movements of the body, is however not typical for human beings. On the contrary, this type of cognitive self-evidence is totally absent once psychic systems are formed, once living beings become human. The main reason for this is that humans, due to their initial helplessness, are from the very beginning dependent on the fellow human being for the interpretation of the stimulus. The interactive patterns of the human infant are from the very beginning marked by the interpretations from and interactions with the other. Psychic structures are not just "pushed" by biological movement processes, they are also "pulled" by the surrounding context in the first place constituted by the fellow human being (cf. GEERARDYN et al. 2002). There will be no psychic structures in the absence of one of the two poles.

Therefore, we need to come to other types of identification, that also rest upon processes of closure, but that witness of an essentially different interactive process, leading to essentially different interaction- and process closures. Apart from the real identification briefly sketched above, there are two other forms of identification that need to be mentioned, the imaginary and the symbolic.

In *imaginary identification*, the "foreign element" is the image in the mirror, taken as representing the own being. Imaginary identification provides the unity of the body via its presentation in the image. LACAN will say that the image is constitutive of the formation of the psychic system (LACAN 1977a). This type of identification was acknowledged also by René Thom and many ethologists, including Lorenz and Eibl-Eibesfeldt, when they stated that the moment the predator chases the prey, the predator *is* the prey.

But this presentation is not fully correct for what concerns human beings. The peculiarity of humans is that they do not restrict themselves to *imaginary identification* as it takes place in animals. If something is taken as an image, standing for one's own being, it is already with the help of the fellow human being. That is why LACAN is talking of the symbolic matrix that precipitates the child in a primordial form: "(…) This jubilant assumption of his specular image by the child at the *infans* stage, still sunk in his motor incapacity and nursling dependence, would seem to exhibit in an exemplary situation the symbolic matrix in which the *I* is precipitated a primordial form, before it is objectified in the dialectic of identification with the other, and before language restores to it, in the universal, its function as a subject."(*Ibidem*: 94).

Already here, the Symbolic enters the picture. It makes us aware of the fact that real, as well as imaginary identification, although in different ways, serve as supports for the constitution of psychic structures, but do not actually constitute the psychic structures themselves. If the psychic system emerges as a structure, then it can only be in as far as it is inscribed in the symbolic order. If the psychological is to be distinguished as a genuinely new organizational layer, with regard to the biological of which it is an emancipation, then it is because of the symbolically constituted structures, or rather, because of the fact that language itself is a structure in which the human infant needs to insert himself. To become human, is to be part of language. As a consequence, the Imaginary is not a genuine organizational layer, it is not organizationally closed in any sense. Moreover, the process of emergence of a psychic structure is not to be understood in terms of developmental stages, with first the Real, then the Imaginary, then the Symbolic identification.

The basic idea of FREUD and LACAN is that humans are born in language — language is for them the most important developmental constraint — and most of the time come to assume the idea that words can represent them, thereby immediately experiencing the inadequacy of words as they can never *fully* represent them. *Symbolic identification* concerns the identification with the signifier (cf. LACAN 1977b). It involves taking a signifier as standing for one's own being in some respect. More precisely, it involves inscribing one's own being in the structure of signifiers, making oneself a part of the chain of signifiers.[8] Whereas LACAN, in the beginning at least, conceived of the signifier as something that is inscribed in a system and which takes on a psychic value purely by virtue of its difference from the other elements (cf. SAUSSURE's interpretation of structure), I believe it is more interesting to attempt to conceive of the psychic structure along biosemiotic lines (which is not incompatible with LACAN's later viewpoints), that is, as a living cohesive system, to be compared much more to a "metabolism" of the signifier than to a static structure of signifiers that are defined purely in terms of positions and relations.

Conclusion

In summary, the way of being in language participates in the dialectics of the living; it quite naturally, though not unambiguously expresses the way of being in the world, in a bodily, spatial and temporal manner. The organic body is reappropriated in language for beings who are produced in and through language. From the moment human beings enter in language, the accessibility of the biological order changes and is coloured by the signifying practices characteristic of language-determined beings.

Real, Imaginary and Symbolic identification cooperate in the construction of a psychic structure. They involve processes of closure in which movements, perception of images and symbolic encounters with fellow human beings take place. The fellow human being and his way of being in language are certainly determinative. It could be interesting to start by considering the fellow human being in terms of perturbations of the biological structure of the organism. The other perturbs in a more or less drastic way the present biological rhythms; he perturbs the various productions of the child by interpreting them, he perturbs the child by being absent or present. Thus we have two interacting structures, one that is in language in a particular way, and one that is not yet in language but is under-interpreted and at the same time over-interpreted. The three forms of identification described here, lead to a progressive differentiation of a self that is inherently mediated by the interpreting other, and that, for that reason, has become in a sense permeable, intrinsically divided between what is pushing and what is pulling, between the internal dynamics of the drive, and the external, global impact of the other, between the biological and the social order.[9] The self is an ever shifting stabilisation of interaction processes. Whereas imaginary identification fixes the meaning, symbolic identification, making a human being into a human being, leads to a recognition of the truly shifting status of the self. Giving meaning to this self remains for ever "in-between".

Author's address

Gertrudis Van de Vijver, Department for Philosophy and Moral Science, Ghent University, Blandijnberg 2, B-9000 Ghent, Belgium. Email: gertrudis.vandevijver@rug.ac.be

Notes

1 Strangely enough, various philosophical approaches have succeeded in missing this point: either by attempting to locate meanings in the things they refer to without acknowledging the "processing" aspect, the *to* somebody (materialism), or by locating meanings in the head or in the mental processing system without acknowledging the "referential" aspect, the *for* something (idealism–formalism–cognitivism).

2 Complexity refers to the number of independent variables required to describe the system; simple systems require only a few variables. Organization refers to the number, scope and dynamics of the relations between the components. For the classification, see Collier/Hooker (1999, pp8ff).

3 A sub-class of these systems are autonomous, anticipative and adaptive (AAA). Autonomy refers to the active independence, a characteristic organizational property, in which responses to signals are interrelated so as "to preserve the organized complexity that underwrites control of that very responsiveness and adaptability" (Collier/Hooker 1999, p5). In that sense, these responses are also adaptive (adaptable and adapted), because they are capable of actively responding to internal and external fluctuations. They are anticipative in that their "capacities imply that their actions anticipate responses that will support those properties.

4 Most authors have *observed* the cohesive, global aspect of living systems, and *inferred* from this that those systems must possess a particular kind of organization. Kant, for instance, in his *Third Critique,* hypothesized that living systems, because they are non-decomposable wholes, must have a circular causal organization, in which everything is simultaneously cause and effect. Cybernetics, first and second, as well as embryology, reintroduced the idea of circularity, and phrased it in the middle of the 20th century in terms of feedback processes. Merleau-Ponty in *The Structure of Behavior* talks again of circular causal processes, and nowadays, it is common to understand these processes in terms of multiple feedback processes or non-linearity. See, for instance, the opinion of Walter Freeman, in describing the workings of the brain: "Linear causality fails most dramatically in studies of the relations between microscopic neurons and the macroscopic populations in which they are embedded. Each neuron acts onto a myriad of others within one to a few synaptic links, and already the returning impact of those others alters its state before it can send another pulse. This hierarchical interaction cannot be reduced to a linear causal chain. (…) particles making up the ensemble simultaneously create a macroscopic state by the very state they have created. A better description (…) is provided by circular causality." (Freeman 1999, pp135–136)

5 The concept of organizational closure is an extension of the idea of an autopoietic system, which is "organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components that: (1) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (2) constitute it (…) as a concrete unity in the space in which they exist by specifying the topological domain of its realization as such" (VARELA 1979, p13).

6 COLLIER continues: "Cohesion is an equivalence relation that partitions a set of dynamical particulars into unified and distinct entities along the lines of John PERRY's unity relation (PERRY 1970)." The relation of unity, as a part–whole relation, refers to the closure of the relations between the components, which implies a global entity to arise. As such, it is different from an identity relation.

7 Therefore, the system needs a kind of infrastructure (MORENO/RUIZ-MIRAZO 1999): an internal organization of processes, which displays energetic as well as informational aspects (BICKHARD 2000). The information concerns the differentiation and switching of the underlying energetic processes.

8 Signifiers can be words, or other kinds of linguistic material. More precisely, they are particular organizations of the (mainly) linguistic material that have had a determining impact in the subject's history, and that function as building blocks for the constitution of the psychic system. For a more detailed account, see BAZAN et al. (2002).

9 This is the way in which I am inclined to interpret LEMKE's idea of a new level that always emerges "in between" two other levels, that reinterprets the variety from the level below *for* the level above. In our case, the psychic system clearly emerges in between the organic and the social level, by interpreting the own body in function of the other (LEMKE 2000).

References

Baldini, F. (1990) Corpo e mente. Progetto di un'antropologia psicanalitica. Thelema. La psicanalisi e i suoi intorni 2.

Bazan, A./Geerardyn, F./Knockaert, V./Van Bunder, D./Van de Vijver, G., (2002) Language as the source of human unconsciousness processes. This issue.

Bickhard, M. H. (2000) Autonomy, function, and representation, CC-AI. The Journal for the Integrated Study of Artificial Intelligence 17(3–4):111–132.

Boudailliez, S (ed.) (1998) De l'identification à l'identité. Special issue of Carrefours, vol. 2: Lille.

Buchler, J. (1955) Philosophical papers of Peirce. Dover Publications: New York.

Christensen, W. D./Hooker, C. A. (2000) Anticipation in autonomous systems: Foundations for a theory of embodied agents. In: Dubois, D. (ed) Proceedings of the Third International Conference on Computing Anticipatory Systems, Casys'99. American Institute of Physics: New York.

Collier, J./Hooker, C. (1999) Complexly organised dynamical systems. Open Systems and Information Dynamics, vol. 6: 241–302.

Collier, J. (in press) Self-organization, individuation and identity. To appear in Revue Internationale de Philosophie.

Dalto, S. (1998) La genèse du sujet dans la perspective du matérialisme scientifique freudien. In: Geerardyn F./Van de Vijver, G. (eds) Aux sources de la psychanalyse. Les premiers écrits de Freud (1877–1900). L'Harmattan: Paris, pp. 263–273.

Freeman, W. J. (1999) How brains make up their minds. Weidenfeld & Nicolson: London.

Freud, S. (1957) Instincts and their vicissitudes. In: Strachey, J. et al. (eds) The standard edition of the complete psychological works of Sigmund Freud, Volume 14. Hogart Press: London, pp. 109–159. Originally published in 1915.

Geerardyn, F./Van de Vijver, G./Knockaert, V./Bazan, A./Van Bunder, D. (2002) How do I know what I think till I hear what I say: On the emergence of consciousness between the biological and the social. This issue.

Grunberger, B./Chasseguet-Smirgel, J. (eds) (1978) L'identification. L'autre, c'est moi. Tchou: Paris.

Hoffmeyer, J. (1996) Signs of meaning in the universe (Translated by B. J. Haveland). Indiana University Press: Bloomington & Indianapolis. Originally published in 1993.

Hoffmeyer, J./Emmeche, C. (eds) (1999) Biosemiotica. Special issue of Semiotica 127(1–4).

Kauffman, S. A. (1993) Origins of order: Self-organization and selection in evolution. Oxford University Press: New York and Oxford

Lacan, J. (1961–62) Le Séminaire. Livre IX. L'identification. Publication hors commerce de l'Association freudienne internationale: Paris.

Lacan, J. (1977a) The seminar. Book IV. The four fundamental concepts of psychoanalysis (Translated by A. Sheridan). The Hogarth Press: London. Originally published in 1964.

Lacan, J. (1977b) The mirror stage as formative of the function of the I as revealed in psychoanalytic experience. In: Ecrits: A selection (Translated from the French by Alan Sheridan). W. W. Norton: New York. Originally published in 1949.

Lemke, J. (2000) Opening up closure: Semiotics across scales. Annals of the New York Academy of Sciences, special issue on Closure: Emergent Organizations and their Dynamics 901:100–112.

Maturana, H. R./Varela, F. J. (1980) Autopoiesis and cognition: The realization of the living. Reidel: Dordrecht.

Moreno Bergareche, A./Ruiz-Mirazo, K. (1999) Metabolism and the problem of its universalization. BioSystems 49:45–61.

Pattee, H. (1995) Evolving self-reference: Matter, symbols, and semantic closure. Communication & Cognition–Artifical Intelligence 12(1–2):9–27.

Rosen, R. (1985) Anticipatory systems, Pergamon Press: Oxford, England.

Rosen, R. (1991) Life itself. Columbia University Press NY.

Salthe, S. N. (1985) Evolving hierarchical systems. Columbia University Press: New York

Van Bunder, D./Knockaert, V./Bazan, A., Van de Vijver, G./Geerardyn, F. Some remarks on the organization of human speech: The unconscious structured as a language. Unpublished manuscript.

Van de Vijver, G. (1997) Signs and Systems. Review of J. Hoffmeyer (1996) Signs of meaning in the universe, Bloomington & Indianapolis, Cybernetics and Human Knowing 4(4):65–68.

Van de Vijver, G. (1998) Emergence et explication. Intellectica 25(2):7–23.

Van de Vijver, G. (1999) Psychic closure: A prerequisite for the recognition of the sign-function? Semiotica 127(1/4):613–631.

Van de Vijver, G. (2000) Identification and psychic closure a dynamic structuralist approach of the psyche. In: Chandler, J./Van de Vijver, G. (eds) Closure: Emergent organizations and their dynamics. Annals of the New York Academy of Sciences 901:1–13.

Van de Vijver, G./Salthe, S./Delpos, M. (eds) (1998) Evolutionary systems. Biological and epistemological perspectives on self-organization and selection. Kluwer Academic Publishers: Dordrecht.

Varela, F. (1979) Principles of biological autonomy. North Holland Elsevier Company: New York.

Zusammenfassungen der Artikel in deutscher Sprache

Paul Bach-y-Rita
„Volume Transmission" und Gehirnplastizität

Aufbauend auf den Studien von CAJAL und SHERRINGTON dominierte in den Neurowissenschaften das Konzept der synaptischen Übertragung. Kommunikationsprozesse innerhalb des Gehirns, welche auf interzellulärer Diffusion diverser Substanzen beruhen (diese wird auch als „volume transmission" [VT] bezeichnet) galten als ausgeschlossen. Der vorliegende Artikel befasst sich mit dieser Art von Informationsübertragung.

VT besteht in der nichtsynaptischen Diffusion von Substanzen durch extrazelluläre Flüssigkeit. Dabei kommt es im Gegensatz zur synaptischen Übertragung nicht zu einer 1:1 Übertragung von Information, sondern zu einer langsamen, interzellulären Übertragung. VT erweist sich dabei als wesentlich energiesparender als die synaptische Übertragung und spielt u.a. eine große Rolle bei Regenerationsvorgängen nach Hirnverletzungen. Zahlreiche Untersuchungen legen nahe, dass synaptische Übertragung und VT sehr häufig kombiniert vorkommen.

Weiters wird in diesem Artikel auch dargestellt in welcher Art die neurowissenschaftlichen „Moden" eine Integration radikal neuer Konzeptionen verzögern.

Elizabeth Ennen
Die Basalganglien und das kognitive Unbewusste

Ausgangspunkt der Überlegungen ist das sog. „habit memory". Zahlreiche neurowissenschaftliche Untersuchungen machen deutlich, dass die Basalganglien das überwiegend unbewusste „habit memory" bilden. Nähere Darstellungen dieses „habit memory" sind dabei insoferne widersprüchlich, als dieses einerseits als nichtrepräsentativ und andererseits als kognitiv–repräsentativ interpretiert wird.

Im vorliegenden Artikel wird von der Existenz multipler Gedächtnissysteme ausgegangen, welche sowohl repräsentationale, wie auch nicht-repräsentationale Speicherformen beinhalten, welche beide am Zustandekommen von Verhalten und komplexen Intelligenzleistungen beteiligt sind.

Stephen William Kercel
Interne Ursachen – bizarre Effekte

Ausgangspunkt ist die schon von SCHRÖDINGER gestellte Frage, inwieweit Konzepte aus den sog. „harten" Natur- und Computerwissenschaften Erklärungswert hinsichtlich Lebensprozessen und auch mentalen Prozessen beinhalten. Schon bei der Untersuchung relativ einfacher Organismen (wie beispielsweise Nematoden), die über ca. 300 Neuronen verfügen und deren Entwicklungsmuster bestens analysiert sind wird deutlich, wie unzureichend die vorhandenen Modelle (beispielsweise das konnektionistische) sind. Noch deutlicher werden diese Unzulänglichkeiten bei komplexeren Strukturen.

Als eine mögliche Ursache dieses Zustandes kann der noch immer vorherrschende CARTESISCHE Dualismus angenommen werden.

Die Arbeit versucht eine Überwindung dieses unbefriedigenden Zustandes mit Hilfe sog. internaler semantischer Systeme. Damit soll eine Perspektive jenseits der sog. „computation" entwickelt werden, die eine Eigenschaft von Gehirnen ins Zentrum rückt, die bei Computern nicht anzutreffen ist.

Leib Litman/Arthur S. Reber
Der zeitliche Verlauf der Konsolidierung impliziten Wissens

Der Beitrag beschäftigt sich mit einer Reihe nach wie vor ungelöster Probleme hinsichtlich der Konsolidierung von implizitem und unbewusstem Wissen. Zahlreiche empirischer Untersuchungen hinsichtlich Lern- und Gedächtnisleistungen legen die Annahme nahe, dass unterschiedliche Formen (implizit/unbewusste versus explizit/kognitiv/bewusste Formen) der Informationsverarbeitung und Repräsentation existieren, die auch in unterschiedlichen neuronalen Systemen lokalisiert sind.

Die hier behandelten Fragestellungen thematisieren vor allem die Konsolidierung und die damit verbundene zeitliche Dimension der Entstehung von implizit gelernten Gedächtnisinhalten im Vergleich zu explizit (bewusst) gelernten Inhalten.

Die in diesem Artikel vorgestellten empirische Untersuchungen legen die Annahme nahe, dass eine Wissenskonsolidierung sowohl bei implizitem wie auch bei explizitem Lernen notwendig ist. Darüber hinaus wird deutlich, dass beim Erwerb impliziten Wissens (im Ggs. zum expliziten Wissen) der Hippocampus keine zentrale Rolle spielt.

A. Minh Nguyen
„Blindsight" und das Unbewusste

Unter „blindsight" („Blindsehen") versteht man die Fähigkeit von blinden Menschen, Objekte in ihrer Umgebung anzuzeigen bzw. ihre Position zu bestimmen. Diese Arbeit geht davon aus dass „Blindsehen" auf unbewusste Erfahrungen zurückzuführen ist. Dabei zeigen „blindsight" Menschen trotz der Aussage, dass sie überhaupt nichts wahrnehmen und nur raten, bei Orientierungs- und Bewegungsaufgaben fast perfekte Leistungen.

Folgende Einwände, welche gegen die Annahme unbewusster Erfahrungen gerichtet sind werden behandelt:

- Die logische Unmöglichkeit der Annahme unbewusster Prozesse.
- Da die „blindsight" Personen von sich behaupten keinerlei visuellen Erfahrungen zu haben und bei diversen Aufgaben bloß zu raten ist die Zuschreibung dieser Erfahrungen bzw. Prozesse aus der Aussenperspektive unzulässig.
- Die Annahme von unbewussten Prozessen und Zuständen welche den Leistungen von „blindsight" Personen zugrunde liegen ist nicht die beste Hypothese um die vorliegenden Daten zu erklären. Es gibt andere und bessere Hypothesen.

Ariane Bazan/Filip Geerardyn/ Veroniek Knockaert/David Van Bunder/ Gertrudis Van de Vijver
Sprache als Quelle unbewusster Prozesse beim Menschen

In diesem Artikel wird ein Konzept zur Erfassung unbewusst determinierten Verhaltens, welches in psychischen Symptomen und Träumen zum Ausdruck kommt präsentiert. Dabei werden einige klinische Beobachtungen innerhalb eine FREUD–LACAN´schen Perspektive operationalisiert. Das menschliche Unbewusste wird dabei als ein linguistisch strukturiertes dynamisches System aufgefasst. Diese psychoanalytischen Überlegungen werden mit aktuellen, neurowissenschaftlichen Erkenntnissen zur Sprache in Beziehung gesetzt.

Menschliche Sprache erscheint dabei auf zwei evolutionär unterschiedliche neurologische Schaltkreise zurückzuführen zu sein, welche teilweise parallel arbeiten. Das ältere System ist dabei subcortical angelegt und vor allem für affektive und phonemische Eigenschaften der Sprache zuständig, während das evolutionär „jüngere", spezifisch menschliche System, neocortical angelegt ist und vor allem semantische Eigenschaften der Sprache bedingt.

Sehr wichtig sind in diesem Zusammenhang die Beziehungen zwischen den emotionalen Erfahrungen und der jeweils spezifischen Sprachdynamik.

Die zentrale Hypothese geht dabei von der Existenz eines linguistischen Unbewussten aus, welches darin ersichtlich wird, dass unausgesprochene Phoneme (z.B. aufgrund von Verdrängungprozessen) sog. „Phantomphoneme" im linguistischen Aktionsraum bedingen. Diese „Phantomphoneme" würden mentale Vorstellungen bedingen, die nun nicht semantisch sondern vor allem phonetisch organisiert sind.

Wilma Bucci
Die Sprache der Emotionen: Eine evolutionäre Perspektive

Menschen und andere Organismen vollziehen permanent eine Fülle informationsverarbeitender Prozesse in unterschiedlichen Sinnesmodalitäten. Als Besonderheit des Menschen kann dabei jene Eigenschaft gesehen werden, die ihn dazu befähigt die Fülle nonverbaler Verarbeitungsprozesse in einem gewissen Ausmaß mit dem Sprachsystem zu verbinden. Dabei zeigt sich jedoch, dass der sprachliche Ausdruck bestimmter Erfahrungen (wie etwa emotionaler Erfahrungen oder die verbale Darstellung des Geschmackes einer bestimmten Weinsorte) beträchtliche Probleme bereitet. Daraus wird ersichtlich, dass Sprache nur einen geringen Teil der inneren Erfahrungen zum Ausdruck bringen kann.

Um Adaptationsprozesse zu gewährleisten müssen diese unterschiedlichen Subsysteme integriert werden. Dabei liegen unterschiedliche Klassifikationen dieser Subsysteme vor. So sieht beispielsweise S. PINKER den menschlichen Geist als ein Integrationsprodukt vielfältiger angeborener, evolutionär bedingter Module.

Als zentrales Problem der vorliegenden Arbeit erscheinen jene Prozesse, die diese unterschiedlichen Subsysteme integrieren. Die in zahlreichen Konzeptionen vertretene Annahme der Sprache als Integrationsinstanz wird hier verworfen und ihr die sog. „multiple code theory“ gegenübergestellt. Demnach vollzieht sich Informationsverarbeitung auf zwei Ebenen – der symbolischen und der subsymbolischen (oder nicht symbolischen), welche durch den sog. „referential process“ bis zu einem gewissen Ausmaß integriert werden.

Robert E. Haskell

Das neue kognitive Unbewusste

Zentralproblem sind die sog. unbewussten semantischen Referenten, d.h. jene Referenten die gemeinsam mit den bewussten, konventionellen Bedeutungsgehalten von Worten existieren.

Methodologisch wird dabei der Versuch unternommen diese unbewussten semantischen Referenten nicht wie bisher mit psychoanalytischen Methoden zu ergründen, sondern diese innerhalb der Kognitionswissenschaften zu untersuchen. Die Ergebnisse legen eine neue Sicht unbewusster Prozesse nahe, deren Operationen systemisch und internal konsistent beschaffen sind. Diese Resultate werden abschließend mit evolutionären Überlegungen in Beziehung gesetzt.

Bence Nanay

Evolutionäre Psychologie und das selektionistische Modell neuronaler Entwicklung: Ein kombinatorischer Ansatz

Evolutionäre Psychologie und selektionistische Theorien neuronaler Entwicklung werden normalerweise als getrennte Bereiche betrachtet, deren Schwerpunkte einerseits phylogenetisch und andererseits ontogenetisch gelagert sind. In diesem Artikel wird der Versuch einer Kombination unternommen um damit auch einige Kritikpunkte (Angeboren/Erworben Dichotomie; Adaptionismus; Modularismus), die an beide Theorien herangetragen werden zu entkräften.

Steven M. Platek

Unbewusste Reaktionen auf kindliche Gesichter: Der Ähnlichkeitseffekt

Diese empirische Arbeit geht von der soziobiologischen Annahme aus, dass das Ausmaß elterlichen Investments in Kinder vom Grad der Gewissheit die eigene Elternschaft betreffend, abhängt. Im Experiment wurde dabei das Ausmaß der Reaktion auf ein kindliches Gesicht in Abhängigkeit von der Ähnlichkeit von Gesichtszügen zwischen VP und Kind untersucht. Die Ähnlichkeit wurde einerseits durch computergestützte Analyse der Gesichtszüge und andererseits durch sozial vermittelte Ähnlichkeitszuschreibung (der VP wurde gesagt, dass Ähnlichkeit mit dem Kind besteht) bestimmt.

Dabei zeigte sich, dass Frauen wie Männer im gleichen Ausmaß auf sozial vermittelte Ähnlichkeitszuschreibung reagierten, während Männer in stärkerem Ausmaß auf die aktuellen Ähnlichkeitszuschreibungen reagierten. Alle Versuchspersonen gaben dabei an in keiner Weise bewusst auf Ähnlichkeiten geachtet zu haben.

David Livingstone Smith

A Breast of Flesh Air: Die Evolution unbewusster verbaler Kommunikation

Das Phänomen der „Selbsttäuschung“ bzw. des „Selbstbetruges“ scheint evolutionär aus der Notwendigkeit heraus entstanden zu sein die eigenen Absichten vor den anderen zu verbergen. Dabei entsteht eine Form unbewusster Kognition, die derart wohl nur beim Menschen existiert. Hier wird die Auffassung vertreten, dass Selbsttäuschung einen starken Selektionsdruck auf die Entstehung komplexer Formen unbewusster sozialer Kognition und Kommunikation darstellte. Menschen überprüfen vielfach unbewusst das Verhalten der Anderen auf Anzeichen von Täuschung. Dabei wird diese Information unbewusst analysiert und das Ergebnis fließt sodann vielfach in die verbale Kommunikation ein, indem dabei unbewusste Bedeutungsgehalte aktiviert werden.

Das Datenmaterial wird dabei vor allem aus psychanalytischen bzw. anders orientierten Therapiesitzungen gewonnen. Eine weitere Datenquellen, welche die Existenz unbewusster sozialer Prozesse deutlich machten waren sog. „T-Gruppen“ (psychotherapeutische Gruppen die aus einen Trainer und einigen Teilnehmern bestehen).

H. John Caulfield

Überlegungen zum Selbstbewusstsein – seine Evolution und Erweiterung auf Artefakte

Ausgangspunkt ist die Frage ob sich ein „Selbst“ bzw. ein „Selbstbewusstsein“ problemlos in Lebewesen mit Bewusstsein entwickeln würde. Ich versuche dabei die Entwicklungslinie aufzuzeigen und dabei deutlich zu machen, daß „Selbstbewusstsein“ sobald dieses einmal entwickelt ist durch die Evolution stark gefördert wird.

Dieser Ansatz basiert auf folgenden Annahmen:

Keine Magie bzw. Wunder; Quantenmechanische Argumente werden nur dort angeführt, wo einfachere, „klassische“ Erklärungen versagen; in Erweiterung biologischer und neurobiologischer Ansätze schließt der hier vorgestellte Ansatz auch nichtbiologische Elemente (Artefakte) ein; es werden nur solche Ansätze akzeptiert, die der menschlichen Alltagserfahrung des „Selbstbewussteins“ nicht widersprechen.

Bislang wurde kein Konzept entwickelt, welches auf den hier vorgestellten Annahmen beruht.

James H. Fetzer

Entwickelndes Bewusstsein: Die zentrale Idee

Die Suche nach Erklärungen für die Evolution des Bewusstseins wird auch als das „hard problem“ bezeichnet. Die damit verbundenen Probleme werden durch Konzepte wie „unbewusst“ und „vorbewusst“ beträchtlich erschwert. Um den Einfluss unbewusster Faktoren auf das Bewusstsein zu verstehen benötigt man Wissen über die Evolution wie auch über das Bewusstsein selbst. Die vorliegende Arbeit unternimmt den Versuch einen theoretischen Rahmen zum Verständnis von Evolution und Bewusstseins zu entwikkeln, wobei die Schwerpunkte auf den genetisch–evolutionären Funktion des Bewusstseins einerseits sowie auf den kulturellen Kontexten andererseits liegen. Dabei wird deutlich, dass das „hard problem“ vielleicht doch nicht ein solches ist.

Menschlicher Geist bzw. mentale Phänomene im Allgemeinen werden dabei als semiotisches System interpretiert. Dieser Ansatz stellt eine plausible Alternative zu den vorherrschenden computer- und sprachzentrierten Ansätzen dar und ermöglicht ein tiefergehendes Verständnis von evolutionären und kognitiven Prozessen.

Temple Gradin

Haben autistische Menschen und Tiere echtes Bewusstsein?

Die weit verbreitete Annahme, dass Bewusstsein von Sprache abhängig sei wird in diesem Beitrag einer Kritik unterzogen. Im Zentrum steht dabei die These, dass die Dominanz visueller Vorstellungen (das Denken in Bildern) – als Eigenschaft autistischer Personen – einen ebenso zentralen bewusstseinsbildenden Faktor darstellt.

Bei autistischen Menschen zeigt sich, dass die Sprache keine notwendige Voraussetzung bei der Bildung von Konzepten und Kategorien darstellt, sondern auch visuelle Eindrücke eine zentrale Rolle spielen und damit auch bewusstseinsbildend wirken.

Autistische Personen sind daher voll bewusst. Verschlossen bleiben ihnen jedoch hoch abstrakte sprachlich formulierte Erfahrungen. Ebenso zeigen sich Defizite im Bereich des sozialen und emotionalen Bewusstseins. Ausgehend von diesen Überlegungen wird die Annahme vertreten, dass Bewusstsein kein einheitliches Phänomen darstellt, sondern vielleicht – ähnlich wie bei Gardners unterschiedlichen Denkformen – unterschiedliche Typen von Bewusstheit existieren, wie beispielsweise ein Angstbewusstsein, ein Schmerzbewusstsein etc.

Filip Geerardyn/Gertrudis Van de Vijver/
Veroniek Knockaert/Ariane Bazan/
David Van Bunder

„Wie weiß ich, was ich denke, bevor ich höre, was ich sage“: Über die Emergenz von Bewusstsein zwischen dem biologischen und dem sozialen Bereich

Dieser Beitrag behandelt vor allem jene Prozesse, die sich (vor allem im Rahmen psychoanalytischer Sitzungen) bei der Bewusstwerdung vorher nicht bewusster Inhalte vollziehen. Als zentrales Problemfeld erscheint dabei der Übergangsbereich zwischen der biologisch – neuronalen Ebene sowie der sozial – sprachlichen Ebene. Das dabei auftauchende sog. „Leib–Seele“ Problem wird bei FREUD in einer nicht dualistischen und nicht reduktionistisch–materialistischen Weise behandelt. Der biologische Gehalt der FREUD'schen Überlegungen wird jedoch im Rahmen der Psychoanalyse häufig missachtet. Unter Weiterführung der Überlegungen FREUDS und in Anlehnung an LACAN wird in dieser Arbeit ein Ansatz entwickelt, der ausgehend von den Begriffen Emergenz und Signifikant (Bedeutungsträger) versucht die Kluft zwischen Neurowissenschaften und Psychoanalyse zu überwinden.

Dori LeCroy

Freuds Konzept der Abwehrmechanismen in evolutionärer Perspektive

Setzt man das FREUD'sche Konzept der Abwehrmechanismen mit dem in der evolutionären Psychologie entwickelten Konzept der „Selbsttäuschung" in Beziehung, so können die Abwehrmechanismen auch als adaptive Verhaltensstrategien interpretiert werden. In dieser Arbeit werden unterschiedliche Formen des Bindungsverhaltens und die im Zusammenhang damit auftretenden Abwehrmechanismen aus dieser Perspektive heraus untersucht. Dabei zeigt sich, dass sowohl die Abwehrmechanismen, wie auch das Unbewusste überhaupt adaptiven Gehalt aufweist.

Gertrudis Van de Vijver/David Van Bunder/
Veroniek Knockaert/Ariane Bazan/
Filip Geerardyn

Die Bedeutung von Geschlossenheit für psychische Systeme im Rahmen einer dynamisch–strukturalistischen Sichtweise

Die sog. „organisatorische Geschlossenheit" ist eine der Grundvoraussetzungen für ein dynamisches System damit dieses stabil, selbsterhaltend und selbstreproduzierend sein kann. Dies gilt ebenso für psychische Systeme, die in der vorliegenden Arbeit aus einer dynamisch–strukturalistischen Perspektive heraus untersucht werden. In Analogie zur „organisatorischen Geschlossenheit" lebender Systeme allgemein, erscheint der Begriff der Identifikation als Schlüssel zum Mechanismus der psychischen Geschlossenheit sowie der Bildung psychischer Strukturen. Dabei werden unterschiedliche Formen von Identifikationsprozessen unterschieden welche an der Ausbildung und Differenzierung des „Selbst" maßgeblichen Anteil haben.