Contents

Impressum

Evolution and Cognition: ISSN: 0938-2623 **Published by:** Konrad Lorenz Institut für Evolutions- und Kognitionsforschung, Adolf-Lorenz-Gasse 2, A-3422 Altenberg/Donau. Tel.: 0043-2242-32390; Fax: 0043-2242-323904; e-mail: sec@kla.univie.ac.at; World Wide Web: http://www.kli.ac.at/ **Editors:** Rupert Riedl, Manfred Wimmer **Layout:** Alexander Riegler **Aim and Scope:** "Evolution and Cognition" is an interdisciplinary forum devoted to all aspects of research on cognition in animals and humans. The major emphasis of the journal is on evolutionary approaches to cognition, reflecting the fact that the cognitive capacities of organisms result from biological evolution. Empirical and theoretical work from both fields, evolutionary and cognitive science, is accepted, but particular attention is paid to interdisciplinary perspectives on the mutual relationship between evolutionary and cognitive processes. Submissions dealing with the significance of cognitive research for the theories of biological and sociocultural evolution are also welcome. "Evolution and Cognition" publishes both original papers and review articles. **Period of Publication:** Semi-annual. **Price:** Annuals subscription rate (2 issues, postage and handling included): Euro 40, US$ 50, SFr 65, GBP 28. Annual subscriptions are assumed to be continued automatically unless subscription orders are cancelled by written information. **Single issue price** (postage and handling included): Euro 25, US$ 30, SFr 35, GBP 18. **Publishing House:** WUV-Universitätsverlag/Vienna University Press, Berggasse 5, A-1090 Wien, Tel.: 0043/1/3105356-0, Fax: 0043/1/3197050. **Bank:** Erste österreichische Spar-Casse, Acct.No. 073-08191 (Bank Code 20111). **Advertising:** Vienna University Press, Berggasse 5, A-1090 Wien. **Supported by Cultural Office of the City of Vienna, Austrian Federal Ministry of Education, Science, and Culture, and the Section Culture and Science of the Lower Austrian State Government.**

Robert Artigiani

The Evolution of Humans and Human Evolution

Abstract

Biological science has made unchallengeable contributions to understanding human origins. However, it does not follow that biological science can explain everything human. Nevertheless, renewed attempts to account for all human attributes as aspects of our genetic heritage and hard wiring are being made. Several of these efforts are ambitious and many are based on sophisticated reasoning and careful experimentation. Despite their quality, this essay argues that striving to reduce culture to biology violates the first principles of an evolving nature. A symmetry-breaking discontinuity between the evolution of humans, which biology maps nicely, and human evolution, the emergence of species-specific attributes like morality and consciousness, is postulated. A distinct explanatory framework based on the self-organization of complex systems is then offered to track the emergence of meaning and selves within a less reductionist scientific paradigm. In this way a bridge hypothesis between biology and culture is offered, for the patterns characteristic of evolution are retained while the content on which these patterns operate is different.

Key words

Complexity, evolution, self, values-ethics-morals (VEMs).

In "The Origin and Goal of History", Karl Jaspers said, "The unity of man in the movement of his metamorphoses is not a static unity of persisting… Man has become man in history through a movement that is not a movement of his natural make-up. As a natural being he is given his make-up in the area open to the play of its variations, as an historical being he reaches out beyond this natural datum" (Jaspers 1953). I quote these lines, from one of the twentieth century's leading Existentialist philosophers, to show how difficult it is to capture the idea that people change over time. Arguing that history matters—that what we value in our humanity emerges over time from our biological origins—Jaspers distinguished between biological and cultural evolution. Like his great contemporary Ernst Cassirer (1960), Jaspers insisted that "culture" must be treated on its own terms and explained according to its own principles.

Recently there has been a vigorous reassertion of the notion that culture cannot be decoupled from biology. In fact, there are currently numerous researchers who are finding the origin of behaviors, like art, and ideas, like religion, in basic biological structures. Going beyond the obvious claim that all things human have biological antecedents, these scholars "reduce" fully developed human attributes like culture and consciousness to biology (Allott 1994). Evolutionary science is used to defend these claims, for they rest on the assertion that what people are and do depends on physiological characteristics first selected by nature almost 3,000,000 years ago. Of course, no one still argues everything we do is biologically determined or "merely" biology (Dunbar/Knight/Powers 1999). But, protests to the contrary notwithstanding, there are claims leaving morality and consciousness, effectively, epiphenomena (Fehr/Gachter 2002).

The alternative is not to argue that people are either culturally determined (Dawson 1996) or infinitely malleable, for there are real biological constraints on what we are and do. Moreover, it is essential to respect the accomplishments of evolutionary science. Its explanation of change through a combination of variation and selection is compelling, and that fundamental model can be applied across a vast range of natural processes. But applications of scientific paradigms outside their original frames of reference should be made carefully, and the mechanical materialism currently being applied to the study of culture is inappropriate. It is, in any case, being replaced by a new

scientific paradigm, which emphasizes processes over things, symmetry-breaks over continuities, and relations over forces (PRIGOGINE/STENGERS 1984). Literal translations of nineteenth century biological concepts, by contrast, make a category error by accounting for what happens on the social level of reality by detailed descriptions and analyses of what exists on the biological level of reality (WRIGHT 1994). Showing how qualitative change takes place, the new scientific paradigm provides a basis for balancing the claims of nature and nurture and understanding cultural evolution in terms close to those of humanists like JASPERS and CASSIRER.

II

The emergence over time of our particular biological species, the "evolution of humans", can be distinguished from "human evolution". In the first case, the subject is rightly understood in terms of natural science. The challenge for scientists interested in the evolution of humans is to understand how a bipedal creature with species-specific characteristics like an exceptionally developed neo-cortex, an opposable thumb, and language evolved. Since what is at stake is a set of physiological attributes, it is reasonable to look for their origins in chemical molecules and mechanical processes. The physiological characteristics of human pelvises, crania, and voice boxes can then be accounted for by describing how complex molecules—genes—evolved by random variations and natural selection. These chemical molecules instruct other molecules so that organic matter forms and acts properly. The potential of various possibilities is captured in individual organisms and tested by environmental selection. Thus, every organism is a hypothesis a species makes about its world.

When DARWIN first suggested the idea of evolution through the environmental selection of random variations the world was scandalized. When Mendel discovered the chemical mechanism through which attributes could be inherited and variations produced, the world was oblivious. No one, apparently not even DARWIN, paid any attention. But since the early twentieth century, as biologists have married selection and genetics into an "evolutionary synthesis", the success of DARWINISM has been overwhelming (MAYR 2001). To be sure, there are some reasonable refinements being made around the edges of the new synthesis. DARWIN may not have understood the sources of order very well, and certainly his disciples have overestimated competition. Nevertheless, for the purposes of this essay, subtle innovations can be ignored, and I will operate with an essentially DARWINIAN model, refined to include "punctuated equilibrium" (ELDREDGE/GOULD 1972) and self-organization (KAUFFMAN 1993).

Since success encourages expanding a paradigm there is nothing surprising about attempts to apply the basic idea that evolution can be understood in terms of chemical processes which are randomly varied and tested against environmental realities to new territories, as any DARWINIST would admit. These pioneering efforts should be encouraged. But they should always be treated skeptically and they should always be undertaken with that spirit of inquiry which is the hallmark of scientific investigation. It is fair to use every successful application of a paradigm as the springboard for its expansion, but when the expansion is undertaken investigators should be alert to the possibility that in new domains new maps may be necessary. A becoming modesty expects no less—populations, after all, can migrate into environments that prove inhospitable.

To at least some of us in the humanities, recent efforts to explain "human evolution" by expanding the paradigm successfully used to explain the evolution of humans violates this seemingly obvious caution (COSMIDES et al. 1992). Enthusiasts bring tools used to explain joints, skulls, and tissues to the challenge of explaining behavior, consciousness, and identity (MITHEN 1996). This strikes me as an example of the old legend about the man who could hammer better than anyone else in the world. His accomplishments were all well and good, and they deserved to be admired. But his habit of seeing everything as a nail often proved counter-productive. Similarly, when it comes to explaining morality and self, the chemical tools of geneticists, the distribution models of mathematicians, and the physical mechanisms of environmentalists often seem to flatten the basic issues to such an extent that their meaning is lost.

The Sociobiological ideas associated with E. O. WILSON (1975) and the selfish genes of Richard DAWKINS (1976) are cases in point. Based on exquisite knowledge of ants, biochemistry, and computer programs, these efforts to explain why humans feel, think, and act the way they do today are framed by strict allegiance to natural science techniques. They are essentially reductionist, arguing that the root of even the most sophisticated behavior or most highly developed attribute can be found in some sort of biological wetware. Thus, mind and self are reduced to brain states and morality to games genes play.

Trying to account for morality and the self, current scientific thinking at least recognizes the existence of attributes clamoring for explanation (MARIJUAN 2001). But it too often then supposes that these distinct phenomena can be accounted for using conventional techniques (DAMASIO 1999). This *is* progress, for traditional Modern science simply denied the causal efficacy of intangible qualities, if not their very existence. For classical Modern scientists, morality and self were ghosts in the machine. In their NEWTONIAN world, morality and self had no real existence. Morality and self may feel real and qualitatively different from instincts and bodies. But, said the Modern materialists, these apparently distinct phenomena were only consequences of mechanisms whose complicated workings had yet to be fully described and quantified. When, said T. H. HUXLEY, we calculate the mechanical equivalent of consciousness, for instance, its status would be fully understood. HUXLEY's intellectual descendants are pursuing a similar agenda, seeking, for instance, to mathematically demonstrate that altruism is actually an effective strategy for passing genes along in certain carefully circumscribed conditions. There are broad, general claims, as well, some of which locate culture and religion in biologically determined behavior (LUMSDEN/WILSON 1981).

A few classic Modern scientists, including HUXLEY, had the grace to recognize that there was a difference between the laws of the jungle and the laws of the boardroom (HUXLEY 1893). HUXLEY rejected the notions that businessmen were the apex of evolution and that they achieved their dominating position in the rugged landscape of corporate finance by being more ruthless, aggressive, and deceptive than their competitors. HUXLEY did not dispute that businessmen were cruel, selfish, and hypocritical. He just denied that their behavior could be justified by an appeal to evolutionary ethics. But his alternative was not much more edifying, for he denied there was a scientific basis for ethics. He claimed that people were ethical, but he believed that ethics developed in opposition to nature. Nature was as DARWIN had described it, but societies were not natural. They were like "gardens", HUXLEY thought, and within the range of their artificial boundaries the biological laws science described were held at bay. We should cherish our artificial values and live by them for as long as possible, HUXLEY said, always remembering that our little gardens were carved out of an infinite universe whose ultimate power and moral opacity could never be forgotten.

III

There are possibilities in contemporary science that suggest we can do better than either HUXLEY or his epigoni. Following, for example, Ilya PRIGOGINE, we can locate certain patterned processes which seem to operate across a wide range of natural phenomena but which are nevertheless not reductionist. PRIGOGINE stated his ideas succinctly by asserting that nature is too rich to be described in a single language (PRIGOGINE 1977, p51). In this thought, among the most beautiful in all of science, he recognized that there are things about chemical reactions for which physics cannot account, and things about biological processes that chemistry cannot explain. He also seems to propose that some ecological phenomena elude explanation in biological terms and to suggest that there would be aspects of human social existence that transcend the explanatory capacity of ecology. Levels of reality differ qualitatively, and we can only respect each of these levels by using a language appropriate to it. To be "scientific", therefore, does not mean to speak only of quantitative, material elements, for all of nature is not made of countable physical objects.

The difference between "the evolution of humans" and "human evolution" represents one such variation in levels. This asymmetry warns against extending biological science to morality and consciousness. Morality and consciousness are real, but they are a level of reality where the rules that apply cannot be reduced to genes. Among these rules are, for instance, the idea of duty that sent firemen at Chernobyl and the World Trade Center to certain doom helping total strangers. The existence of duty does not qualify the ability of biological science to explain the "evolution of humans". If we wish to know how the physiological entities characterized by bipedalism, precise hand–eye coordination, and speech came into being, then genetic variation and natural selection are exactly the right way to do it—at least so far as we know. But what the evolution of humans explains should be clear—it is the emergence of a population of organisms equipped with attributes for surviving in a wide range of environments by individually processing local flows in electro–chemical terms. Biological humans have a highly evolved sense for learning about, evaluating, and reacting to various physical conditions. We sense in terms of color and shape, sound and taste, feeling and odor. And we react to what we sense in terms of pleasure and pain, sickness and health, fear and fury. But it is a far cry from admiring the skill

with which genetic variation and natural selection explain biological humans to claiming that ethics are genetic or self-awareness is a brain state.

Of course, were there no fundamental genetic programs or physiological states there would be no morals or selves. But that is like saying that if there were no physical and chemical basis there would be no eyes or lungs. Yet no chemist observing ion transmissions, etc. before the emergence of sighted land-dwelling animals could have predicted eyes and lungs. The rules of perception and oxidation would not have been sufficient to project so boldly. After the event, of course, the chemical bases of eyes and lungs are traceable. But the beauty of DARWINISM is that, because it does not determine emergent characteristics, it need not reduce levels of reality to their predecessors. Random mutations, fluctuating boundary conditions, and altered energy flows create the surprises that punctuated equilibrium maps. Morals and selves are as surprising as eyes and lungs, and as irreducible to the laws of chemistry and physics.

Eyes and lungs are staples of the biological arsenal, for they are the sorts of organs opponents of evolution point to as proofs there must be a designer involved in creation. Eyes are such marvelously precise and detailed instruments for seeing that, say the anti-evolutionists, they must have been designed for that purpose. But if eyes evolved gradually from very inferior instruments for processing light, there would probably not have been sufficient competitive advantage in the first examples to warrant their selection. So, say the anti-evolutionists, God must have designed and built them when His time was ripe.

Such arguments once carried weight. But we have since learned that the original eye was not an organ for seeing a all. It was a heat sensor developed by organisms living in shallow water. Heat is a form of energy, like light, so it turns out that the development of a heat sensor laid a base for the later development of a light-processor. But nature did not develop in a straight line, building incremental change upon incremental change over eons of time. That is the way things would work should events be under the command of a Supreme Scientist, Philosopher, or God. But, said Nobel Laureate Francois JACOB, nature is closer to an engineer or even a tinkerer (JACOB 1977). Evolution is not, therefore, some grand rational scheme worked out in advance and driven toward an intentional goal. It is a haphazard process in which whatever is available is snatched up and put to some practical use to solve an immediate problem. Evolution takes what has developed for one purpose and redefines it by putting it in a new context. Lungs are very complicated devices for breathing air but they did not originate with that purpose in the mind of anyone or anything. Nor did their initial form determine their final role. Lungs were initially flotation bladders used for stabilizing sea-dwellers. When some of those creatures found themselves living at least part of the time in dry air, random variations in the rules for making flotation bladders turned them into lungs.

Something similar, I suspect, occurs with phenomena like morals and selves, so it would be tragic, in explaining them, to simply repeat the error anti-evolutionists make explaining eyes and lungs. Practicing the science anti-evolutionists attacked, we should never try to find evolved attributes at the beginning of a process. To find a moral gene, a neuronal self, or a God area in the brain is to suppose that our current capacities can be traced back to sources deterministically acting to produce those capacities. What is worse, insisting that all discussions of human attributes be reduced to biology is tantamount to acting as if the evolutionary process reached its apogee with biologically modern humans and stopped. Doing so not only repeats the errors of anti-evolutionists, it is contrary to the very idea of evolution.

Evolution maps a dynamic nature in which qualitative changes occur. To look beyond biology in accounting for our humanity is in the spirit of evolutionary science. Until biological researches locate the appropriate genes and brain areas, therefore changing categories to accommodate cultural evolution seems reasonable. After all, words being cheaper than genes, I am only claiming that, in culture, nature has found a faster way to explore possibilities. Evolutionary theory supports such explorations. Properly used, it provides a bridge hypothesis that allows us to use the concepts of science without limiting ourselves to the content of physics, chemistry, or biology. Thus, morality and selves can be seen to "emerge" rather than as having to be explained in terms of mutations and brain states.

To be sure, the evolution of humans developed some genetic propensity toward, among other things, sharing and caring. Had that not been the case the big brained babies prematurely delivered through birth canals shrunken to accommodate bipedalism would never have survived long enough to reproduce. Similarly, distinct parts of the brain are especially active when decisions about behaviors with immediate personal consequences, like gathering food and sharing it, are being made, just as there

are parts of the brain engaged with speech. Since planning actions and letting others feed from the same trough involve simple behaviors and organic tissue, they probably do have a genetic base.

But none of these points means that morals and consciousness are ultimately explainable in terms of developed sensors and brain modules. Morals are not simply sensations that are pleasant or painful, nor is consciousness located in a particular part of brain tissue or in the brain operating as a whole (FISCHER 1990). Morals and consciousness are as qualitatively different from mechanical actions and tissues as eyes are from heat sensors and lungs from flotation bladders. They all "emerged" in much the same way, so the patterned process explaining how eyes and lungs evolved should be tested against problems like morality and mind. But we should not expect that the same organic categories used to explain the origins of organs and tissues fit the emergence of morals and selves: Nature is too rich to be explained in a single language. Even if the same conceptual model for evolution generally applies across a wide range of phenomena, what is being explained at different levels of reality can vary profoundly. We need, therefore, to stop talking about human evolution using the same categories and substances that worked so well in explaining the evolution of humans (TATTERSALL 1998). Instead we should try to understand the historical process by which people evolved in the human sense. We do that by telling the story of how biologically human beings became conscious, moral, and individuated in a language that respects morality and thought.

IV

But we need not join with T. H. HUXLEY in supposing that the human world of morality and awareness violates the norms of nature. HUXLEY may have been a leading proponent of DARWINISM, but he was, after all, a devotee of the Modern scientific paradigm. NEWTON was still, for him, the standard for scientific explanation. The only problem was that NEWTON's paradigm described a world without history, a world that literally could not evolve.

Things moved in NEWTON's world, but they did not change. Hence they could be moved in the opposite direction and arrive back at their precise initial conditions. Time, in such conditions, was "only an illusion". But in contemporary science, which rests more on thermodynamics than physics, time is real and has a direction. In time, things do not merely move; they change. Events are irreversible and they have a history. So, for scientists of the Modern era from 1650 to 1950, if there was a realm where ethics rather than competition existed, that realm was not natural and could not be accounted for scientifically. For postmodern scientists, on the other hand, nature is replete with symmetry-breaking changes. It is quite possible, therefore, to see ethics emerge in time by following natural processes.

The Modern definition of science was too rigid to accommodate intangible realities like ethics. Assuming that the external world was exclusively material and independently existing, disciples of BACON and GALILEO took their role to be collecting information about nature by observation and experimentation. Today we realize that nature is not external to us but something of which we are entirely part. Since we are part of nature and we are alive, it must not be exclusively a material reality, for how the material is organized makes a huge difference in what it is and does. The chemicals in organisms are no different than chemicals in dead material bodies. But organisms are radically different from bodies, and the things living matter can do are radically different from the things dead matter does.

Different though they may be, living bodies are no less real than dead ones! Contemporary science can deal with these differences, for it concentrates on mapping how nature works rather than describing what nature is. Ultimate reality is no longer the dead stuff outside our brains but the interactive processes in which we participate and through which nature creates itself. Describing what nature is led Moderns to their ahistorical universe of undeveloping movements. Concentrating on how nature works is leading contemporary scientists to an understanding of how nature changes. Thus, this new science is more realistic than NEWTON's. Our world has a history, and any science that cannot account for its evolution has limited itself to describing laboratory idealizations rather than natural realities.

The vehicle for change in nature is interaction (PRIGOGINE 1997). Ironically, interaction is the same process that troubled the founders of quantum theory BOHR, HEISENBERG, and BORN, for it leads to the transformation of components in the same way that observation led to the appearance of "phenomena". Contemporary science sees this process in a positive light, whereas for the Copenhagen Interpretation it was negative. Copenhagen, aspiring like good Moderns to know what the nature out there is, felt betrayed by its method when the act of observation was shown to change the observed. Doing science, Copenhagen concluded, meant interacting with na-

ture and losing information. Hence scientists could no longer say what the observed was in itself (RAE 1986). Obliged to report pointer readings embedded in apparatus, they faced the "limits of science" (SULLIVAN 1949), which left quantum theory an end of the road hypothesis (POPPER 1982).

Making observations may change what exists. But that does not mean science has reached the end of its road. Rather, reaching the limits of Modern science might just expand the frontiers of understanding. Because an observation is an interaction between an element and a flow, understanding what science does gives clues to how nature works. There are flows —gradients—throughout nature, and they interact with objects of all sorts, after all. Instruments similarly submit an observed entity to some sort of energy flow. Since scientists are part of the nature they are trying to understand from the inside, realizing that what happens in science is happening in nature should not be surprising. But if it is reasonable to suppose nature is actively involved in observing itself, there is no need to suppose that nature actually seeks knowledge. It does, however, make sense to suppose that when a flow interacts with an entity in nature, the same sort of changes occur that plagued quantum physics. Nature, like observed elements in laboratories, is changed by these interactions. But the implications are now reversed: Where Modern science saw interaction losing information, contemporary science sees nature creating information through interaction. It is the mechanism by which nature evolves.

There is another famous aspect of quantum theory that we should consider, namely the problem of objectivity. BOHR defended science against the charge that changing nature made results subjective. Instead, he argued that scientists repeating observations using the same apparatus in similar circumstances would achieve the same results: their findings were as "objective" as BACON's. They might not have been "real", but they were public, repeatable results about which everyone agreed. It was impossible to deny, however, that quantum results were laboratory "phenomena" and that they were "embedded" in the observational apparatus. For contemporary science, these embedded phenomena become keys to understanding evolutionary processes. The new levels of reality that emerge over time now appear to be results embedded in systems that "self-organize" when thermodynamic flows perturb elements and create information.

Self-organization is natural, for it tracks the spontaneous emergence of order out of chaos. No Designer or external Source intervenes to create organization, for systems organize themselves. All that is required is that elements in an energy flow are bounded in some way. Nothing has to happen, of course. Elements can simply be incapable of interacting, or the arena in which they interact can be sized in such a way that elements interacting in weak flows lose contact with each other quickly. Most obviously, interacting elements might not organize because energy flows within narrow boundaries are so fierce that organization shatters as fast as interactions occur. But if the elements interact in just the right way with just the right frequency in just the right space, a coherent system might self-organize, defining "right" in all these cases as it does so. Given the multitude of possibilities, there is no reason to suppose particular systems had to self-organize. But once systems have self-organized they display logic and coherence, balance and harmony that looks as designed as eyes or lungs. Changes in components caused by interactions are now embedded in thermodynamic flows the way phenomena are embedded in apparatus. The changes stored in self-organized systems, being observable by others, are "objective". They are nature's evolved realities.

Once systems self-organize they display attributes that may not even be implied in their component parts. Romantic couples provide an obvious example. If we look at each partner separately, as, for instance, the therapists MASTERS and JOHNSON did, we will find two sexually motivated individuals seeking pleasure by friction. We can explain all that biologically—or even chemically. But sometimes two people interact in ways that unite them into higher-level systems, e.g., "romantic couples", and when that happens each is transformed. Their sexual motivation and delight in the pleasure each gives the other are not lost, fortunately, but something has been added to the union—"love"—which is not in their chemistry and may not always even be pleasurable.

As MASTERS and JOHNSON found after decades of research, their measures of individuals told them nothing about love. The reason is simple: love does not exist in separate individuals. It is an attribute of romantic couples and operates top-down from the system-level to change what the partners are. Most obviously, the biological strategies each may have followed in nature seeking and approving breeding partners gives way to a commitment demanding exclusive loyalty and fidelity within the pairing. In the best of circumstances, each member of a romantic couple accepts the obligation owed to the couple gladly, for, in love, they have gained an attribute that

was not in their genes. We can model the emergence of romantic couples following the example of self-organization typical of the rest of nature, but we must use the language appropriate to their level of reality. Measuring chemical flows or pulse rates explains sex; to understand a loving relation takes entirely different terms.

V

Love can stand, metaphorically, for all that is qualitatively distinct in an evolving nature. It exemplifies the irreducible something that emerges when a system self-organizes. Self-organization is the vehicle for evolution—when a system self-organizes it stores information created by interacting elements. The way interaction transforms elements is obvious in the case of a romantic couple. As they meet, flirt, talk, share intimacies, and relate, each is trying out a repertoire of behaviors the other might find attractive. A kind of selection and co-evolution occurs. Any encouragement tends to exaggerate the favored behavior and discourage alternatives previously associated with the particular person. Thus, the person in a romantic couple is as different from the separate individual as, say, carbon in a cell is from carbon in coal or diamond. More excitingly, what each finds attractive is at least in part a consequence of the effects his or her own actions have had on the other. Thus, in changing the other, each is changed in ways neither would expect. If the changes each makes resonate with the expectations raised in the other, a bonding may occur which is so surprising Italian tradition calls it the "Thunderbolt". Each partner is physiologically the same as the person who entered the relationship—before and after DNA tests would be identical. But behaviorally, lovers are completely different from sex partners.

Classic examples of such interaction-based transformations abound in the literature (AUGROS/STANCIU 1987). A great favorite is lichen, which is a distinct life-form created by the interaction of algae and fungi. Once the lichen emerges, the algae and fungi are lost. If the lichen is later analyzed, the algae and fungi are restored. But BLAKE was right—"to dissect is to murder". If we isolate the algae and fungi, the lichen no longer exists. Something similar happens when large numbers of people interact in mutually transforming ways and social systems self-organize. Morals and selves may then be created by interactions and stored in self-organized social systems. If we look for these emergent characteristics in the biological heritages of the people who make societies up, however, we are looking at the wrong level.

It is tempting to look for morals and selves in the biological components that are the base for societies because we can see individual people. That is because we are biologically equipped to perceive biological persons. So it is "natural" for scientists impressed by biology's success in explaining the evolution of organisms to suppose that they can understand societies by reducing them to their visible components. We are then left with the rather comic image of bourgeois gentlemen, doubtlessly dressed like Hobbes and Locke, negotiating with each other about how to form a society. But there was more transformation than negotiation. The independent selves Modernity imagined creating society are actually the products of its self-organization (MEAD 1962). Scientific observers misdirected their explanatory efforts, for it is as hard for them to detect social systems using their biologically prepared receptors as it is to perceive the four dimensional space–time continuum in which our universe is embedded. Science has accepted the latter challenge; it is time to accept the former, as well.

The strongest clue to the existence of social wholes, however, may be the very qualitative attributes we are trying to explain. Assuming that nature does nothing in vain, we can ask the same question that evolutionary psychologists and philosophers ask: what is the selective utility of attributes like morals and selves in nature? Even if nature is not the ruthlessly competitive realm VICTORIAN businessmen envisioned, it must be a world in which people more-or-less exclusively attend to what goes on at the moment in their immediate surroundings. People need to sense that world, understand its potential threats and opportunities, and react to them quickly. Variation and natural selection equipped people very nicely for doing that, and the information necessary for reproducing those attributes was economically stored in DNA. People sensed the world around them, experienced pleasures and pains, and fought or fled, fed or coupled accordingly. Almost all their children developed the same skills and behaviors in a few years.

An economical nature would not seem very likely to have invested much time and energy in laying the bases for—let alone developing the full potential of—morals and selves. To ruminate over long-term consequences, wonder what it all meant, and determine one's relative worth, in natural circumstances, would seem more harmful than helpful. But nature did do things like develop heat sensors and flotation

bladders, which later developed, unplanned and unexpectedly, into eyes and lungs. I think something similar happened with our consciousness and consciences. They rest on physiological characteristics like large brains with highly developed neo-cortexes and voice boxes. But those characteristics were developing as responses to randomly generated solutions to strictly physical problems. It turns out that bigger brains offering better correlations between hand and eye are effective in a variety of ways, while voice communications makes it possible for us to hunt, gather, and fight together in groups even more effectively than, say, LORENZ's beloved wolves. Still, it is a category error to equate being smarter at solving practical problems with exploring the meaning of life or expressing meaning through artistic and religious discussions of our human identity.

If we answer questions about what life means in biological terms like pleasure, pain, and reproductive rates, the results are about as relevant as the famous "42" issuing from the cosmic computer in Douglas ADAMS's "Hitchhiker's Guide To The Galaxy". Like numbers, pleasures, pains, and reproductive rates can be excruciatingly precise. But they are not always relevant. They tell us how well the biological level of reality is working when the question became why the biological realm exists. Since such questions would not arise in a purely organic world, it makes no sense to answer them in purely organic terms. If we want to know how and why humans evolved the behavior of asking after the meaning of life, we need to look at circumstances in which questions about meaning and purpose might reasonably arise. But we would then recognize that the sorts of issues evolutionary psychology and Sociobiology describe are functions of these altered circumstances. The new issues arose on the basis of evolved biological attributes, for they depend on our ability to map the world around us. If, however, by interacting people changed each other and stored information created about themselves in self-organized social systems, the world observed would be different (KANT 1963). Mapping a different world requires a new language.

VI

No detailed reprise of the prehistoric processes by which human societies emerged is necessary for present purposes. Nor need we commit ourselves to any particular interpretation of the disputed facts. We can simply take the most obvious of possibilities and suppose that evolved humans, able to adapt to various environments, multiplied. In their initial conditions, most biologically equipped individuals would have been environmentally fit. But their very fitness would mean that increased numbers stressed environments. Sooner or later, and time after time, small bands of humans grazing in virginal territories would have prospered and multiplied. But when their numbers multiplied, they found that competition with other equally fit neighbors for plants and animals could get fierce. Then it would become extremely difficult for even the sharpest eyed, fleetest-footed, and most cunning individual physical competitor to survive—there would simply be too few ripe fruits and vegetables and too little game left to sustain the swollen population.

In such circumstances, bands reached bifurcation points. Their members could either fight each other to oblivion, settle down and work harder and more cooperatively, or break up into smaller units and wander off in search of more accessible resources. Since people biologically indistinguishable from us show up in the archeological record about 100,000 years ago and civilizations do not appear until about 6,000 years ago, it seems fair to suppose our ancestors most often decided against both mutual destruction or extra work and organization. But as they fragmented and migrated, many important developments occurred. Tool-making became ever more sophisticated, and burial rituals marked the first signs of social stability. No doubt the role of language expanded exponentially, and the first evidence of 'art' appeared.

Much has been made of decorative art, which may first have been performed on the bodies of individuals. It has traditionally been supposed that decorations reflect the joy of creation and the expression of Self. If this is the case, then we can expect to find a biological base for decoration and hypothesize that its appearance resulted from some as yet unconfirmed mutation (PFEIFFER 1982; KLEIN/EDGAR 2002). While comfortably reductionist, there is no need to postulate a genetic mutation. Rather, we can easily see decoration as an indicator that an evolutionary punctuation occurred when social systems self-organized in response to resource pressures (STINER et al. 1999). There are several strategies for explaining societal self-organization (BOGUCKI 1999), some quite systemically sophisticated. But the easiest way to imagine societies emerging is in response to population pressure. For the sake of convenience, that model will be used here. The critical point is that, when environments were desiccated, fitness was measured by the ability to cooperate. But this shift

need not depend on a biological mutation. Nor should it, for hunting faster moving, more agile animals and gathering smaller, less nourishing aquatic creatures could successfully be hunted using team work. Biology might have found genes for improving the speed and accuracy of individual hunters. But if there were insufficient resources available that would not have helped. Nor would there have been time for biological processes to find effective behavior-guiding genes through trial and error.

But if people started to work together they partly decoupled themselves from "natural selection" as competition shifted to groups (SOLTIS/BOYD/RICHERSON 1995). It then became essential to distinguish group members from outsiders. Genetic variations in physiognomy were not available, but body decorations substituted nicely (STEINER/KUHN 2001). Thus, decorations are not expressions of inherent biological inclinations to know the self but are "social skins" (TURNER 1980). Decorations record the self-organization of new social realities where cultural rather than genetic rules obtained. In a world where evolution has not stopped, the rules of a system actually matter more than the stuff the system is made of (ZIMMER 2002). Operating under rules written in altered environments, components of self-organized systems—human or otherwise—acquire new attributes. When band peoples settled into villages, for instance, polygymous men, who worked no harder than necessary and shared what they produced, became husbands, who labored dependably and hoarder their surpluses (FLANNERY 1972). Moreover, archaeologists are now demonstrating that the reasons for organizing more complex social systems are themselves social, rather than biological (HAYDEN 1992; STARK 1986).

But people did not instantly and dramatically change their natural way of life. In many places boundary conditions were too fluid for complex systems to self-organize. Similarly, energy flows from gathering and hunting would often have been too low to sustain increased complexity. In any case, very loose social systems seem to have predominated until people began moving into river valleys (COULBORN 1969). In the river valleys, boundary conditions changed, resources multiplied, and behavioral revolutions occurred—at least among those bands that malaria did not immediately select against. Freshly watered, these territories rewarded the hard work of clearing and draining with unmatched bounties. But rising resources meant population would grow. For a while, the consequences of population growth would have gone unnoticed, since in valleys like the Nile floods carry silt deposits that restore fertility. Thus, even though population rose, bands were not driven to bifurcation points where fragmentation and wandering were options. Meanwhile, making the valleys habitable had taken a huge investment. Settlement made them ever more effective attractors. It did not take long for nomads to notice the increased bounties and tamed environments, either. Acting like people always had, nomads meandered into agricultural settlements and gathered what was available. But the locals whose work had produced these bounties would have been reluctant to see their fruits carted off. So fortifications were raised, which bound valley-dwellers permanently to each other and certain places (CARNIERO 1970). Interactions among themselves led to the self-organization of complex societies in river valleys, and interactions between nomadic raiders and those societies led to their rapid streamlining.

The change in behavior does not require any new theory. It seems to mimic standard biological patterns for explaining speciation through "budding" and isolation. Unintentionally, valley-dwellers had placed their rudimentary societies in previously avoided environments, where social variations developed exponentially. Tying themselves to varied behaviors and relationships, they had turned swamps and wildernesses into gardens by hard work. Their successes led to increased populations, which no longer split from the original band and wandered off. So even though floods regularly restored the soil, each new generation of settlers had to support an ever larger population by refining and inventing new behaviors.

There are many ways to increase productivity, and valley-dwellers doubtlessly tried them all. But those who survived tended to have specialized their labor. Specialization is a protean development, for it does more than increase individual productivity. It increases interdependence, for specialized individuals spend so much of their time making things others need that they cannot meet needs of their own. Farmers labor in the fields all day, which makes them vulnerable to raiders, weather, and time. For farmers, to survive, soldiers must defend farmers, weavers clothe them, and artisans supply their tools.

Many must have been the times when members of these specialized groups yearned to do what their ancestors had done—or to do nothing at all. But that was increasingly selected against, for while people were meeting their own needs in altered circumstances a social system was self-organizing around them. Acting together, people had created wholes

greater than the sums of their parts. They had built irrigation dams and canals, warehouses, temples, and walls; they had invented writing; and now they were obeying laws. They survived by cooperating. But cooperation meant that people were no longer able to whimsically follow whatever urges or inclinations struck them individually. Whimsical behaviors now threatened the network of mutually reinforcing relationships that sustained them all. Organized societies were now solving problems individuals could no longer solve for themselves. Preserving these solutions societies stored information as different from individual instincts as eyes and lungs are from heat sensors and flotation bladders.

VII

No doubt the evolved capacity to live in groups demonstrated by our simian cousins laid the bases for civilized societies. But this genetic basis for behavior is not the whole story. Genes can vary without affecting fitness (KIMURA 1987), and it is perfectly reasonable to suppose that behavior can vary to some degree without genetic mutations. The relations between genes, environments, and behaviors is neither 1:1 nor strictly deterministic. Besides, as VON NEUMANN (1963) pointed out, much of the information used to construct an organism is stored in structures external to the genes. For that reason, it is sensible to think of three strands of information interacting to create living systems rather than the two characterizing the double helix (LEWONTIN 2000). The environment contributes significantly to the outcome, and there is nothing about the logic of the genes that constrains the logic of its external world. Genes and environment co-evolve. So it is not unlikely that biologically stable human populations facing similarly stressed environments combined into social systems that altered the definition of fitness. Once societies self-organized nature acted directly on them and only on individuals through the mediation of social structures.

But while it was societies that self-organized, the particular characteristics of the individuals alive at the moments when social systems emerged was critical. Therefore different skills and relationships were favored at different times in different places. Thus, the collective responses of groups differed, perhaps only slightly but nevertheless. Since there are feedbacks between individuals and societies and societies and environments, the relations defining each group were nonlinear. As a result, even slight distinctions in societies, over time, could have different survival possibilities. Moreover, variations in the structures of social systems could be amplified to such a degree that they would eventually be at least partly incommensurable. The languages, arts, and religions characteristic of different groups would separate their human populations. Such variations cannot be explained by the same genetic mutation in human organisms, of course.

Once social systems self-organized, a level of reality qualitatively different from the biological emerges. Understanding this level takes new tools, so, for instance, analyses of sociability based on "group selection" are not applicable. We should not look for selection processes acting on individuals that somehow produce results benefiting groups. Nor must we abandon basic evolutionary principles. Rather, we should try to understand societies as emergent entities operating on a new scale. These, the societies, are the population of individuals where environmental selection applies. Societies must be defined in their terms, not in terms of the biological organisms from which societies self-organized.

Since no two societies are exactly the same, biological models of variation and selection should apply to social as well as biological evolution. Thus, as organizations subject to selection, the information structuring societies can be tested in a manner analogous to the way nature tests DNA. Variations structuring societies in different ways can provide advantages similar to those enjoyed by differently shaped or colored organisms. We could even claim, I suppose, that societies are simply machines invented to perpetuate socially structuring information. We need not go so far, however. But we should at least try to shift our scientific focus from the fitness criteria applicable to biological organisms and examine what is at stake on the societal level. In societies, whose boundaries and operations partially decouple people from nature, information guiding behavior rather than physiology comes into play.

It took many centuries for humankind to figure out how to guide behaviors in social systems. Resolving the problem was accomplished by a process of experiment as blind as biological evolution. We should not suppose that prehistoric people sat down and looked forward in time to anticipate the consequences of their various possible actions, opting for those the cleverest people saw would profit them (HUMPHREY 1993). Supposing prehistoric people had the ability to anticipate the global consequences of local actions attributes to them the same mental tools we have. Were that the case, there would be a morality gene and an ethics area in the brain—along,

apparently, with override capacities. Since none have been found, changes in wetware offer no solution to the puzzle. Perhaps the software operating human brains does. In that case we are looking not for the genes producing morals and selves but for the social equivalent of DNA, the information mapping social systems.

Suggesting that human societies wrote programs which transformed brain behavior in fundamental, qualitative ways must seem radical. The idea is actually conservative. It explains changes in people by reference to changes in their environment. The thesis is that when people mapped an altered environment they altered the way their brains operated. To make the maps, however, required no new gene or biologically evolved area. The capacity to map their environments emerged much earlier in simpler lifeforms than our own. Experiments have shown that animals like rats will mentally map an area and thereafter use that mental representation to navigate in it. TOLMAN (1948) called these mental representations "cognitive maps". Neuroscientists understand them well, and our ancestors used this inherited ability to represent the social world.

In self-organized social systems our ancestors were, however, remembering what they experienced in common rather than what they had learned individually (HALBWAS 1992; CONNERTON 1989). They had, therefore, to represent a world of relationships rather than things and actions. Moreover, members of social systems had to map relationships that were not directly perceivable using the sensory equipment biological evolution had developed. Farmers had to know what soldiers, artisans, and clerks were doing even when vast distances or long periods separated them. Since all these specialists were members of the same systems, however, what any of them did would affect the systems' over-all states. So it was quite easy for farmers being overrun by nomads to guess that the soldiers dedicated to protecting a territory had failed. The soldiers may have failed because artisans did not deliver adequate weapons, and that failure might have resulted from clerks miscalculating needs. But whatever the cause, farmers could tell from their experiences whether systemmates far away and earlier in time had behaved as expected.

We can sympathize with those others, because they, too, faced the daunting task of knowing what challenges people they could not sense would face. To know what other members of social systems are doing and needing we have to map the *systems* to which they and we belong. Mapping interactive sets of relationships requires no new cognitive capacity. But now the environments mapped have the capacity to translate what individuals do locally into system-level effects. It is this translation of the local into the global, of what individuals do to what communities experience, that breaks symmetry with biology and produces new cognitive capacities.

The same kinds of processes are at work even as new levels self-organize, however, for flows through systems also provide selection mechanisms. Now, however, selection criteria are social rather than natural. Experience—i.e., selection operating on the system level—teaches people which actions are beneficial and which are harmful, which tend to stabilize social systems and which tend to destabilize them. It is, of course, not enough to teach people after the fact that what they have done was useful or dangerous. People must be able to anticipate the effects of their actions and, in the process, increase the probability of acting in ways that will be stabilizing rather than destabilizing. In these new circumstances behaviors rather than physiological attributes are selected for or against.

Cognitive mapping provides the vehicle for projecting global effects and anticipating local consequences. But projecting what an entire social system is likely to do because of how individual actions affect people remote in space and time requires a very special kind of map. That map is not a particular arrangement of neural nerves, for it is no more a matter of what is happening to an individual than love is limited to one party in a romantic couple. Love is about the couple, and it is the name we give to the ability of the couple to operate top-down on the individual partners. The name we use to symbolize the power of a society to constrain the behavior of individuals and bind them to a collective goal is "morality".

Morals are very special kinds of symbols, for they store information about what people have in common rather than what they know independently about their immediate environments. The first medium available for storing collective information was probably more elementary. I think it was dances and, later, rituals (TURNER 1986). Dances and rituals taught people what particular actions were expected of them in the interest of collective stability by actually putting individuals in physical relationships with other network members and exciting them to repeat the desired behavior. When the dance or ritual was over, the actions were simply continued—albeit in more practical forms. Of course, systems cannot survive in a dynamic nature if they only

know one set of behaviors. So for societies to survive their members had to be able to shift from, say, being farmers to being soldiers. This change recorded the collective knowledge that the social environment contained threats from other groups which had to be met with organized violence. Dances and rituals met this test by engaging individuals acting as farmers, setting them into a repeated rhythmic motion, creating an ecstatic state, and finally transporting those people into behaviors appropriate to warring. Passing individuals through the ecstatic state, social experience appears seamless to its members—they literally dance themselves from one collective capacity into another (LEACH 1976).

Dances and rituals store information about social systems rather than individuals. Information about what people have in common is stored in their collective activities rather than their genes or organs. Performing dances and rituals, in turn, actually constitutes social systems. When children learn to perform these activities, which they do by actually dancing and acting out ritual routines, they are learning how to be members of particular societies. Dances and rituals teach them the predictable, regularized behaviors through which individuals' problems can be collectively solved. Thus societies are examples of non-biological systems writing programs for their own replication.

But dances and rituals have at least one serious drawback: they are time and memory consuming. Dances may take hours, and remembering the various rituals uses lots of neurons. So there are consequently only a few states that ritualistic systems can access, and transitions from one to another tend to be slow. Societies operated by dances and rituals would be near-to-equilibrium and relatively simple. For societies to access many environments and make adjustments rapidly, a more efficient medium for storing and communicating social information was necessary. The obvious alternative to physical action is verbal representation. Language was available, thanks to the evolution of humans, and by exchanging information verbally, people learned to interact and become parts of societies self-organized in thermodynamic flows. Thus, the first step toward "human evolution" was possible thanks to biology—people could talk about their world.

But by talking people changed their world. Instead of acting separately and producing, at most, an additive effect on the environment, language let people work together on their world and produce results far more significant than the sum of their individual labors. This may be why humans proved fitter than Neanderthals, for language made correlating behaviors more effective than individual actions, even by bigger brained creatures, were. Dance and ritual, legend and myth then had to map the world language built, and, describing it, language was, in effect, describing itself. The symmetry-breaking result was the invention of "symbols". Symbols are maps, like brain states, dances, or rituals. But symbols reference the effects of things or actions, not the things or actions themselves. In particular, symbols reference the global effects of local actions. They capture the translations of individual deeds into system states, as when farmers get needed tools and peaceful circumstances because artisans and soldiers were dutiful. In a word, symbols map the *meaning* of actions (BRUNER 1990). For that reason, morals are maps of meaning, which could not have been stored genetically before complex systems existed. Following maps recalling collective experience, people transcended their biological origins, saw themselves from the perspective of others, and became self-conscious (LUCKMANN 1967; SCHWARTZ/LUCKMANN 1989).

VIII

Maps of meaning are stored outside human brains, in "external symbolic storage systems" (DONALD 1991). They are symbols stored externally because they do not map what individuals experience but what individuals do to collectives. Maps of meaning describe the global effects of individual actions. Meaning, in moral as linguistic terms, is translation. Words "mean" the other words into which they are translated, as morals "mean" the global effects of local actions. Thus, if a theory of meaning, in linguistic terms, is as Umberto ECO says, "a theory of the practice of using a language" (ECO/SANTANBROGIO/VIOLI 1988, p4), a theory of meaning in moral terms is a theory of the experience of living in societies. Thus, maps of meaning could not be part of our biological heritage, which would permit them to be stored in our genes and tissues, because the realities maps of meaning reference did not exist until after we evolved biologically. Meaning itself, as a moral category, could not arise before social systems self-organized. Ancient myths realize that, which is why Genesis describes an Edenic state of nature as preceding knowledge of good and evil.

Originally, no doubt, words translated some sort of objects or actions in the natural world. To account for them we need the biological development of voice boxes, tongues, larynxes, and brains with

Broca's area. Having this evolved capacity permitted our ancestors to interact and produce the circumstances in which societies self-organized. But once those societies existed, the world to be mapped by translations into sounds was very different from the one for which humans had been originally selected. Societies created worlds of networks and feedbacks layered over the physical world of rivers and rocks, plants and animals, competitors and breeding partners. These networks were sustained or fluctuated by individual actions, and the effects of those actions were thus translated into consequences experienced by all other members of the systems. Just as individuals in nature needed to predict whether a vegetation would nourish or poison, a game animal would water at dawn or dusk, farmers, soldiers, and artisans needed to know whether their interactive networks would be sustained or endangered by their actions.

Symbolic representations of how social systems react to various stimuli—i.e., to possible individual actions—maps of meaning are like periscopes. They allow individuals, in a manner of speaking, to rise above their local behavioral space and glimpse system wholes. But, like genes, maps of meaning do more than represent their environment. They stimulate actions calculated to reproduce the environment they reference. DAWKINS uses the term "meme" to describe the behaviors that replicate societies. Although meme is a wonderful term, DAWKINS's original use was vague. So I shall refer to symbols constituting maps of meaning as Values, Ethics and Morals (VEMs). Morals map end states, which are valued positively or negatively depending on whether they stabilize or destabilize social systems. Ethics are the rules people learn to get to positively valued end-states and avoid negatively valued ones.

As symbols, VEMs can be present in many brains at once. This capacity accounts for their social selection, for VEMs make ever more massively distributive processing possible. Programs for computing solutions to survival problems posed to populations living in circumstances where genetic endowments no longer prove adequate, VEMS allow social systems to solve many problems at once. The state of social systems at any given time is the solution a population has worked out for generating and processing suitable resource flows. We have seen that generating and processing flows requires people to specialize their behaviors. These specialized behaviors, which are the set of actions permitting others to correlate behavior in familiar flows, are "Social Roles". Social Roles are the improbable behaviors known to produce desired meanings. VEMs are scripts for playing Social Roles; they replicate societies by influencing individual perceptions and actions. VEMs are social analogs of DNA. When people perceive the same realities, evaluate their significance in analogous ways, and respond to their sensations by doing what others expect, social systems are constituted and replicated.

Since replicating VEM-guided behaviors stabilizes the environments in which societies are embedded, predictable flows are released. Those flows reward and punish individuals for the skill with which they played Social Roles. By rewarding and punishing, societies capture biologically developed traits for unexpected purposes in the same tinkering manner JACOB said characterized evolution generally. But now what people are and do is being shaped—not determined—by the network of interactive relations mapped by VEMs. Performing duties that are individually destructive is one consequence of VEMs. Another is hypocritical behavior, for people can conceal private motives, using the VEM shaped perceptions of others to trick them into thinking actions which benefit individual agents help the whole. But moral sacrifice and moral deception are only using VEMs, which are "emergent phenomena" in a literal sense. Mapping a new level of reality, VEMs cannot be reduced to the biological platform from which they emerged. To understand them we need the languages proper to social rather than biological science. But to track the emergence of that language we need only the contemporary scientific paradigm. VEMs are as natural as living organisms, even if they are not reducible to them.

The most obvious example, as DAWKINS said, is God, whose origin biological reductionists have regularly discussed. Whether God's existence can be biologically demonstrated (NEWBERRY/D'AQUILI 2002) is not our problem. But where the idea of God comes from is. Biological reductionists answer the question by seeing God as a manifestation of our need to explain the world causally (BERING 2001), which is fulfilled by a change in brain architecture (MITHEN 1996). Alternatively, we can entertain the possibility that He, She, or It is a "socially constructed reality" (BERGER/LUCKMANN 1990) and clarify what is referenced. God, said the first historian to locate His roots, is a "sense of awe" (OTTO 1958). To me, awe maps the new world of social systems, for it perfectly symbolizes the unexpected experience of the flows on which everyone depends that are released by collective actions involving unseen others. A transcendent, supernatural "spirit" that predates and survives people, knows more than they do, and

exercises life and death power, God could reference the human experience of life in complex social systems *neatly*. Representing that experience, the symbol "God"maps the "semantic loop" (PATTEE 1995) whose closing constitutes societal self-organization (LUHMANN 1995).

The name our ancestors originally gave society, God's initial singularity reflected the relatively simple systems being referenced (ARMSTRONG 1994). Monotheism gave way to polytheism as specialization increased the number of Social Roles crying out for representation and legitimation. Of course, there are limits on the number of Gods which societies can remember, as there were limits on the number of rituals they can perform. Perhaps monotheism revived as an especially austere abstraction because people changing Social Roles needed a symbol able to transcend widely varying experiences. God as a universal, all-knowing and all-powerful spirit supplied the glue rebinding complex societies by beliefs rather than rituals. Each member of a complex society could find the reinforcement needed for his or her particular situation in a transcendent God. Of course, no God is utterly without attributes, and the God societies embrace powerfully affects human actions as they map the human condition. Judging from the turmoil caused by interactions between societies, the symbol sometime seems to work too well. In any case, peoples have developed in different directions under the influence of different Gods.

IX

Living in a world enriched by created information, biological humans gradually changed. They did not go from being unconscious brutes and amoral beasts to urbane sophisticates overnight. Nor was the transition smooth or easy, and it certainly is not complete. But when their biological natures found themselves caught up in self-organized social systems, people began to change in profound ways. Over the period of the last 10,000 or so years people have acquired attributes which we now take to be part of our basic biological condition. Among these are the selves (MAUSS 1985; DUMONT 1986) biological scientists struggle to locate in genes and brain states.

If we suppose people embedded in cultures are different from what they would be in "nature", then it makes sense to search for selves in the opposition between nature and culture. Then the long tradition of lamenting lost innocence and gaining bad habits is less mysterious: our perception of the differences between nature and culture is painful. Whether it is Genesis lamenting the lost Eden or Hesiod condemning the Age of Iron, throughout history people have complained about the strain between instinct and duty, spontaneity and calculation, leisure and labor, equality and hierarchy. In a certain sense, all of these can be summed as the curse of awareness. Awareness ought not be a curse if it originated in biology. There might be psychic strains arising from biological self-awareness, which would probably be no more agitating than, say, sexual frustration. But we were biologically evolved to know the world around us, in great detail and with exquisite precision. Bringing that finely developed sense to bear on our own existence is downright threatening. Even in evolved social environments, in other words, the selective advantage of consciousness is expensive.

That we know ourselves depends more on our social situations than our biological potential. In societies, which are interactive networks, we perceive evidence of our own existence. In a society shaped by the consequences of our actions, when we look at the world we see ourselves (GOULDNER/PETERSON 1962). That would probably be enough to make us jump. But feedback from social systems tells us how what we have done affected other people. It tells us how our actions translate into other people's experiences. Thus, the "selves" we discover in societies are not part of our biological baggage. Selves are social constructions symbolizing the "meaning" of our lives. In civilizations like the Egyptian, selves were originally symbolized as alter ego or companion. In advanced societies, like our own, the Self becomes something like the "Secret Sharer" in Joseph CONRAD's story. The Self becomes almost tangible, and its projected adventures teach us the limits of acceptable behavior. The Self is not the mark of sin but the badge of membership.

We become aware of ourselves when we learn that Others are looking at us. Since natural selection would have favored flight or fight responses to the stares of others, it is no wonder that the emergence of Self was experienced as a threat. In the early days, it was probably not much of a threat, for the links between people in primitive societies would have been few and the behaviors associated with Social Roles fairly natural. With increasing complexity, however, the connections each member has with others and the frequency of interactions multiply. Moreover, complex societies specialize Social Roles, so behaviors become more and more improbable and demand more and more practiced skills. Specialized Roles multiply opportunities for errors and con-

nections provide media for communicating collective appraisals. Finally, complex societies constantly adapt to shifts in environmental circumstances, and every shift alters the relations defining Social Roles. Assuming that we become aware by the experience of other people's observations, the more often societies recalibrate their internal relations in response to environmental shifts, the more often its members will be brought face to face with their Creator and made aware of their own existence (WEBER 1958). The human Self is a classic example of created information.

Of course, when societies and their environments are stable, there is little variation in feedback messages at the several network nodes. Since information reduces uncertainty, stable societies store information about their environments but communicate little information about themselves to their human components. So it is possible for people to live in social systems with only a vague sense of themselves. Since people needed to know more about the groups to which they belong than themselves, CASSIRER (1944) rightly argued that consciousness of the world precedes awareness of the self. Through rituals and myths stable societies also provide devices for soothing anxious brows. In effect, people can be assimilated into relatively simple and/or stable societies so effectively that they seem at equilibrium. There is no news about the Self being forwarded from the system, awareness is sublimated, and the individual becomes one with the community. When the difference between nature and culture is slight, the stress of self-awareness is minimal. No wonder the Chinese sage was thankful not to live in interesting times.

It is in transition eras that people are self-conscious. Either systems are disintegrating and people are learning that practiced behaviors have become dysfunctional, or new systems are self-organizing and people are learning their Roles through the school of hard knocks. In either case, a heightened sense of Self results from multiple observations of individuals by their systems, which are, remember, telling the individuals that they, the individuals, exist and evaluating their worth in terms of a moral identity shadowing their every move. When individuals realize that their survival depends on properly fitting in to their social system, self-awareness becomes uncomfortably associated with actual threats to existence. That is no doubt why, when people become self-conscious in eras of transition, they typically equate regaining social stability with restoring a lost sense of unity and Oneness.

JASPERS located the origin of consciousness in just such circumstances. In his "Axial Age", from 800 to 200 BCE, JASPERS found the source of nearly all major religious and spiritual traditions. All the great thinkers whose ideas are the base of human reasoning—CONFUCIUS, ZOROASTER, BUDDHA, ISAIAH, SOCRATES—lived during this period. And all of them, JASPERS says, became conscious of the Self in one form or another. It is generally accepted by historians that this period, in the several locations where these great thinkers lived, was also a period of rapid growth. Civilization had long existed, but trade was expanding, new Social Roles were being developed, market economies were emerging, and social organizations were being stretched and tested to find means of accommodating new Roles and relations. In other words, people were interacting at increasing rates and new ways, seeking to find behavioral arrangements that would be socially selected.

The strain was awful, and JASPERS was right to admire the creative geniuses who first articulated the experience as consciousness and morality. But he might have looked a little closer at these "Great Renouncers", for most of them were saying that the burdens of consciousness were best escaped. The desirability of earlier, simpler ways of life, when rites were faithfully practiced and people operated at equilibrium with their social worlds, were idealized. Alternatively, the anxiety of living with experience-induced self-awareness was symbolized as alienation and exile. To escape it, people were coached in techniques for losing themselves in states of spiritual bliss that transcended the experiences of everyday life. Even SOCRATES, who believed "the unexamined life was not worth living", played the new Social Role of philosopher in order to show others how to fit in. He was followed by a disciple who advocated "utopias" where everyone would learn their station and its duties, change would stop, and self-consciousness moderate. Certainly poor AUGUSTINE, haunted by the fearsome image of his sinful Self isolated from God, hungered only for a state in which awareness would be lost. In all these cases the experience of the Self was new and dreaded. Looking beyond an Apocalyptic moment toward a restored world of pastoral peace and political unity became increasingly popular.

X

Confronting the gap between nature and society was frightening, but self-awareness proved an effective aid to adaptation. Self-knowledge became a new type cognitive map, and its actual function, as op-

posed to its apparent existence, was depicted by SCHRÖDINGER in "Mind and Matter". Providing an insight more recent biological reductionists might have followed, SCHRÖDINGER wrote that "The reason why our sentient, percipient and thinking ego is met nowhere within our scientific world picture… [is] because it is itself that world picture" (SCHRÖDINGER 1944, p138). The Self as a cognitive map, in other words, is not a representation of the world in which we operate but of the strategies for fitting into that world. Probing environments using organisms, whose fitness cannot be measured for years, would not have been fast enough to invent the Self. Nor would there have been any reason to select for Self genes before the social environment favoring them emerged. Selves as strategies for surviving in the transformed world of social systems, however, could have been conjured out of words and symbols much more quickly. It is much easier to accept an "inner eye" (HUMPHREY 1993) as a metaphor for fitting into social systems than as a genetic mutation meeting the challenge of natural selection.

In complex societies people needed to fit themselves into social systems far more dynamic than nature, and they did it by using knowledge of how previous actions had affected societal networks to construct models of themselves in terms of alternative behavioral possibilities. These models could be fast-forwarded to test alternate strategies. Information about current conditions could be fed into representations of past behaviors. Knowing from VEMs how societies were stabilized or destabilized by different actions, individuals could use representations of themselves to anticipate whether their choices would be rewarded or punished. In complex societies sensitive to even minor shifts in environmental circumstances, frequent projections would have to be made, so our Secret Sharers became regular companions. Eventually, they were reified into actually existing identities, which biological reductionists have been trying to locate ever since.

It was through thinking with others, which the Medieval term "consciousness" implies, that we learned to think about ourselves. But it was not until there were socially constructed defenses and reinforcements that Selves became psychologically viable. It was at the end of another transition period, this one in the 16th century as the Middle Ages disintegrated, that Selves reappeared. Judging from the interest individuals like the goldsmith CELLINI took in their singularity, the roots of a consciousness better able to tolerate self-awareness were probably well established. But the terrible efforts made by religious reformers to restore traditional relations indicate the strain of self-awareness was still threatening. It was only through interactions establishing new institutions, like the nation-state and private property, that resources became available to sustain and protect a heightened sense of Self. Nations for the first time rewarded individuals who created wealth, while private property offered physical defenses for Selves previously lacking. Of course, until new VEMs were articulated, defined, and embraced the process was confusing. But when VEMs for how individuals could cultivate a sense of Self and still fit into the larger systems on which survival depended were finally embraced, a new era in civilization was born. We should not delude ourselves into thinking these current states are biologically given, or even that our attributes were fixed by natural selection in conditions existing 100,000 years ago. It seems more likely that the attributes we take to be species-specific are socially evolved (TAYLOR 1989).

The evolution of morals, Selves, and societies thus go hand in hand. Such parallels usually suggest a causal relation, and it is not hard to derive one here. Social evolution clearly follows the basic patterns of nature. It shows an increase in complexity associated with increased rates of external entropy production. So there is no conscious guiding principle or Designer involved. Nor is any particular outcome fated. The Second Law is simply being obeyed universally by evolving complexity locally, for people in self-organized societies increase the rate at which external entropy is produced. Since there are many paths to increased rates of external entropy production, there is no particular reason for self-congratulation among the Modern, advanced societies. There is not even any reason to congratulate the individuals whose initiatives and explorations triggered the processes by which more complex systems self-organized.

Most of the heroes of Modernization, like Martin LUTHER, intended outcomes quite different from those achieved. So the driver was not individuals whose heroic vision had seen a new and better future. More probably the evolutionary process arrived at Selves the way it stumbled over morals. As societies became more complex they needed more detailed knowledge about their environments and quicker responses to environmental variations. In the first civilizations, the scale at which environments were read was classes or "estates". Classes and estates read environments at such a broad level that they did not learn much. However, the systems to which they reported were stable and could wait upon centralized leadership to determine policy top-down. But as ag-

ricultural civilizations gave way to industrialized ones, societies moved so far from equilibrium that they were in constant jeopardy. Slight environmental perturbations or internal fluctuations, if unattended, could drive societies through cascades of bifurcations into chaos.

Highly differentiated individuals could monitor the environment at a fine-toothed level, and they were willing to do so when economic rewards commensurate with risks were forthcoming. But it was not enough to stimulate individuals to read the environment. They also had to be empowered to act on what they discovered. Far-from-equilibrium societies cannot wait for centralized authority to act before potentially disastrous bifurcations occur. They must operate bottom-up to a large degree, which means individuals must be inner-directed", democratic institutions are essential, and central authorities must be restrained. Popular, constitutional societies are thus as natural as eyes and lungs—and, like eyes and lungs, liberal democracies evolved from rudimentary forms initially playing quite different roles. But in the process of locating institutions like nations, private property, and constitutions, people redefine themselves as much as do the lovers making up a romantic couple or the algae and fungi forming lichen. It is those redefined people we now encounter, whether in the street or in our laboratories. To understand them we need not, as DILTHEY (1962) thought, isolate the study of human evolution from natural science. Thanks to the new paradigm, we can ground our understanding of humankind in nature. All we need do is respect historical experience as much as biological inheritances.

Author's address

Robert Artigiani, History Department, U.S. Naval Academy, Annapolis, Maryland 21402, USA. Email: artigian@usna.edu

Acknowledgments

Frequent and detailed conversations with Penelope DONNELLY and David ISMAY contributed to this essay. Manfred WIMMER suggested the topic and was kind enough to offer scholarly advice. Alicia JUARRERO read drafts and made valuable comments.

References

Allott, R. (1994) The biological basis of poetry. Journal Of Social And Biological Structures 14(4):455–71.

Armstrong, K. (1994) A history of God. Knopf: New York.

Augros, R./Stanciu, G. (1987) The new biology. New Science Library: New York.

Berger, P. L./Luckmann, T. (1990) The social construction of reality. Anchor Books: New York. Originally published in 1966.

Bering, J. (2001) Are chimpanzees mere existentialists? Evolution & Cognition 7(2):126–133.

Bogucki, P. (1999) The origins of human society. Blackwell: London.

Bruner, J. (1990) Acts of meaning. Harvard University: Cambridge MA.

Carniero, R. L. (1970) A theory of the state. Science 169:733–38.

Cassirer, E. (1944) An essay on man. Yale University: New Haven.

Cassirer, E. (1960) The logic of the humanities. Yale University: New Haven.

Connerton, P. (1989) How societies remember. Cambridge University: Cambridge.

Cosmides, L./Tooby, J./Barkow, J. H. (eds.) (1992) The Adapted Mind. Oxford University Press: New York.

Coulborn, R. (1969) The origin of civilized societies. Princeton University: Princeton NJ.

Damasio, A. R. (1999) How the brain creates the mind. Scientific American 281/6:112–117.

Dawkins, R. (1976) The selfish gene. Oxford University: New York.

Dawson, D. (1996) The origins of war: Biological and anthropological theories. History And Theory 35(1):1–28.

Dilthey, W. (1962) Meaning in History: W. Dilthey's Thoughts on History and Society. Edited and introduced by H. P. Rickman. Harper: New York.

Donald, M. (1991) The origins of the modern mind. Harvard: Cambridge MA.

Dunbar, R./Knight, C./Powers, C. (eds) (YEAR) The evolution of culture. Rutgers University.

Dumont, L. (1986) Essays on individualism. University of Chicago: Chicago. Originally published in 1983.

Eco, U./Santanbrogio, M./Violi, P. (eds.) (1988) Meaning and mental representation. Indiana University: Bloomington IN.

Eldredge. N./Gould, S. J. (1972) Punctuated equilibria. In: Schupf, T. J. M./Thomas, J. M. (eds) Models in paleobiology. Freeman: San Francisco, pp. 82–115.

Fehr, F./Gachter, S. (2002) Altruistic punishment. Nature (Jan 10) 415:137–41.

Flannery, K. V. (1972) The origins of the village as a settlement type. In: Ucko, P. J./Tringham, R./Dimbleby, G. W. (eds) Man, settlement and urbanism. Schenkman: Cam-

bridge, pp. 25–53.
Fischer, R. (1990) Why the mind is not in the head. Diogenes 141:3–32.
Gouldner, A./Peterson, R. A. (1962) Technology and the moral order. Bobbs-Merrill: Indianapolis.
Halbwas, M. (1992) On Collective Memory. (Edited and translated by L. A. Coser). Chicago University: Chicago. German original appeared in 1941.
Hayden, B. (1992) Models Of Domestification. In: Gebauer, A. B./Price, T. D. (eds) Transitions to agriculture in prehistory. Prehistory Press: Madison, pp. 11–19.
Humphrey, N. (1993) The inner eye. Vintage: London.
Huxley, T. H. (1893) Evolution and ethics. In: Collected Essays Vol IX. Macmillan: London, pp. 1–116.
Jacob, F. (1977) Evolution and tinkering. Science 196:1161–1166.
Jaspers, K. (1949) The origin and goal of history (Translated by M. Bullock). Yale University: New Haven.
Kant, I. (1963) Idea for a universal history from a cosmopolitan point of view. In: Beck, L. W. (ed) Kant On History. Macmillan: New York, pp. 17–26. German original appeared in 1784.
Kauffman, S. (1993) The origins of order. Oxford University: New York.
Kimura, M. (1987) A stochastic model of compensatory neutral evolution. In: Kimura, M./Kallianpur, G./Hida, T. (eds) Stochastic Methods In Biology. Springer-Verlag: New York, pp. 2–19.
Klein, R. G./Edgar, B. (2002) The dawn of creativity. Freeman: San Francisco.
Leach, E. (1976) Culture and communication. Cambridge University: Cambridge.
Lewontin, R. (2000) The triple helix. Harvard University: Cambridge.
Luhmann, N. (1995) Social systems. (Translated by J. Bednarz). Stanford University: Stanford.
Luckmann, T. (1967) The invisible religion. Macmillan: New York.
Lumsden, C. J./Wilson, E. O. (1981) Genes, mind and culture. Harvard University: Cambridge MA.
Marijuan, P. (ed) (2001) Cajal and consciousness. New York Academy Of Sciences: New York.
Mauss, M. (1985) A category of the human mind. In: Carrithers, M./Callas, S./Luckes, S. (eds) The category of the person. Cambridge University: Cambridge UK, pp. 3–25. Originally published in 1938.
Mayr, E. (2001) What evolution is. Basic Books: New York.
Mead, G. H. M. (1962) Mind, self, and society. University of Chicago: Chicago. Originally published in 1934.
Mithen, S. (1996) The prehistory of the mind. Thames and Hudson: London.
Newberry, A./D'Aquili, E. (2002) Why God won't go away. Ballantine: New York.
Otto, R. (1958) The idea of the holy (Translated by J. W. Harvey) Oxford University: New York.
Pattee, H. (1995) Evolving self-reference: Matter, symbols and semantic closure. Communication and Cognition—Artificial Intelligence 12(1–2):9–28.
Pfeiffer, J. E. (1982) The creative explosion: An inquiry into the origins of art and religion. Harper and Row: New York.
Popper, K. R. (1982) Quantum theory and the schism in physics. Rowman and Littlefield: Totowa NJ.
Prigogine, I. (1997) The end of certainty. Free Press: New York.
Prigogine, I. (1980) From being to becoming. Freeman: San Francisco.
Prigogine, I./Stengers, I. (1984) Order out of chaos. Bantam Books: New York.
Rae, A. (1986) Quantum physics: Illusion or reality. Cambridge University: Cambridge UK.
Schrödinger, E. (1944) What is life and mind and matter. Cambridge University: Cambridge UK.
Schwartz, A./Luckmann, T. (1989) Structure of the lifeworld (Translated by R. M. Zaner and H. T. Engelhardt). Northwestern University: Evanston IL. Originally published in 1973.
Soltis, J./Boyd, R./Richerson, P. J. (1995) Can group-functional behaviors evolve by cultural group selection? Current Anthropology 36(3):473–494.
Stark, B. (1986) Origins of food production in the new world. In: Meltzer, D./Fowler, D./Sabloff, J. (eds.) American archaeology past and future. Smithsonian Institute: Washingtion DC, pp. 277–321.
Stiner, M. C./Munro, N. D./Surovell, T. A./Tchernov, E./Bar-Yosef, O. (1999) Paleolithic population growth pulses evidenced by small animal exploitation. Science 283(5399): 190–194.
Steiner, M. C./Kuhn, S. L. (2001) Ornaments of the earliest upper paleolithic. Proceedings of the National Academy of Sciences (US) 98(13):7641–7646.
Sullivan, J. W. N. (1949) The limitations of science. Mentor: New York. Originally published in 1933.
Tattersall, I. (1998) Becoming human. Harcourt Brace: New York.
Taylor, C. (1989) Sources of the self. Harvard: Cambridge Mass.
Tolman, E. C. (1948) Cognitive maps in men and rats. Psychological Review 55:189–208.
Turner, T. (1980) The social skin. In: Cherfes, J./Lewis, R. (eds) Not work alone. Templesmith: London, pp. 112–142.
Turner, V. (1986) The anthropology of performance. PAJ: New York.
Von Neumann, J. (1963) The general and logical theory of automata. In: Taub, A. H. (ed) Collected Works V. Pergamon: Oxford, pp. 288–328.
Weber. M. (1958) Protestant ethic and the spirit of capitalism. Scribners: New York.
Wilson, E. O. (1975) Sociobiology. Harvard University: Cambridge MA.
Wright, R. (1994) The moral animal. Pantheon: New York.
Zimmer, C. (2002) Is this chip educable? New York Times Book Review March 10:25.

Elisabeth Fivaz-Depeursinge

Emotion and Cognition in the First Year of Life

On Triangulation between Infant, Mother and Father

Abstract

The discovery of the "competent infant" in the early seventies marked a turning point in the knowledge of early development. It showed that affect and cognition are closely interlaced, particularly in secondary intersubjective communication between mother and baby about inanimate objects. A new turning point is emerging as evidence is accumulating on early imitation and triangular interactions of infant with father and mother, revealing a primary form of intersubjective communication in the early months of life.

Key words

Affect, infancy, social cognition, triangular communication.

The discovery of the "competent infant" in the early seventies marked a turning point in the knowledge of early development. For instance, it was then demonstrated that there is continuity in perception from foetal to post-natal stages (Pacteau 1999) and that the neonate differentiates between inanimate objects—they afford exploration—and animate objects—they afford affective synchrony (Butterworth 1995). Likewise, infant learning is motivated and affect-laden (Papousek/Papousek/Koester 1986). It was also shown that young infants actively regulate their attention and affects as they dialogue with their mother; they learn turn-taking and other rules of conversations with affect signals standing in for verbal content (Stern 1985). Thus affect and cognition are closely interlaced already in early infancy.

This interlacing is most visible in the handling of social relationships. At the end of the first year "secondary intersubjectivity" (Trevarthen/Hubley 1978) emerges. It is characterized by intentionality and referential communication. Intentionality is defined as "choosing one's goals and attentional foci for monitoring progress toward goals" (Tomasello/Call 1997, p405) and referential communication as communicating about an object (Bruner 1978). This agenda for social cognition is well established. It slightly precedes Piaget's timetable for physical cognition. In contrast, the evidence concerning the earliest months of infancy is sharply at odds with the notions of indifferentiation, adualism and symbiosis of Piaget's, Baldwin's or Freud's times. This period is characterized by "primary intersubjectivity", a form of communication based on direct, unmediated perception and action (Trevarthen 1984).

A case at hand is the issue of three-person interactions. Recent evidence indicates that young infants engage not only in two-person, but also in three-person interactions; they are attentive to two other persons and have a primitive understanding of how these two persons react to each other. This domain had not been studied until recently. Researchers had been asking the infant whether she handled "triadic interactions", meaning infant and mother interacting about an inanimate object (Bakeman/Adamson 1984). They had discovered that by the end of the first year, the infant coordinated her attention with her mother on objects and deliberately shared her affects of pleasure, frustration or uncertainty with her mother regarding these objects (Stern 1985). However, researchers had not asked the infant whether she handled "triangular interactions", namely *infant and person interacting with or about another person*. This is surprising since, after all, infants are born to a couple and are destined to grow and live in multiperson relationships. One reason for this neglect is that dominant theories in the field (psychoanalysis and attachment) had imprinted the dyadic mother–infant model as *the* matrix of development. Another reason is that as three-person interactions are more complex, their study posed considerable technical and methodological challenges (Parke/Power/Gottman 1979; Barret/Hinde 1988).

But paradigms to study family as well as non-family triangular interactions are now available. New data are emerging that are in need of replication, but are promising. They enlighten the question of affect and cognition in a new way. It is of note that triangular interactions involve four different ways of interacting as a threesome: not only the single "three-together" where all partners are actively engaged, but also the three different "two-plus-ones" where two partners are actively engaged and the third one is in a third-party position. Handling this relational system comes down to negotiating each of the four configurations considered as interchangeable components within a composite system (FIVAZ-DEPEURSINGE/CORBOZ-WARNERY 1999).

The Lausanne Trilogue Play (LTP for short) was precisely designed to observe families as they proceed through these four configurations: 1) mother and baby playing, father as third party (two-plus-one); 2) father and baby playing, mother as third party (two-plus-one); 3) both parents playing with the baby (three-together); 4) father and mother talking together and baby as third party (two-plus-one). The natural goal of trilogue play being to playfully share positive affects, this task primarily targets socio–affective interactions. Technically, the setting allows for recording of the three partners facing each other in a triangular formation, with a sufficient resolution to allow for microanalytical coding of body and facial movement (see FIVAZ-DEPEURSINGE/FRASCAROLO/CORBOZ-WARNERY 1996).

It is important to remember that non-verbal communication is spontaneous, implicit and mostly non-conscious, especially in infancy when moral and cultural rules have not yet made their call for social adaptation.

In this paper, I review data establishing secondary intersubjectivity and introduce new evidence demonstrating primary intersubjectivity.

Secondary intersubjectivity

Coordination of attention

It is established that toward the end of the first year, infants coordinate their attention with their mother on a referent inanimate object. A prototype of this competence is finger pointing. As she points to an object, the infant trusts her mother to understand that she, the baby, has a goal in mind, like getting the object, and that she can share this goal with her mother. Somewhat later, she will follow and then direct her mother's attention (CARPENTER/NAGELL/TOMASELLO 1988). This is a much researched domain for it is an important step in the development of the infant's theory of mind and access to verbal language.

Coordination of affect

However, sharing a focus of attention cannot possibly be devoid of affect, even if it is only interest. It is known from the study of triadic interactions that as an infant becomes enchanted in playing with a toy in the context of an interaction with her mother, she will at some point look up from the object, share her pleasure with mother and then resume playing. This is called *affect sharing*. It is considered as an intentional stance for sharing affect (KASARI/SIGMAN/MUNDY/YIRMYA 1990).

This affect sharing proceeding from an intentional stance is well illustrated in the following observation in the LTP (Lausanne Trilogue Play). A baby is playing with mother whereas father is third party. They take turns at tapping on the seat and the baby looks up at mother, saying: "heu". Mother laughs. Then baby turns to father, saying "heu". Father, wary of strictly keeping to his third party role, does not respond. The baby is surprised. This is not in style with the father's usual behavior. Baby continues addressing father. Father looks up at the ceiling, as if distracted, and this is even more perplexing for the baby, who continues to challenge him. Finally, father cannot resist any more; he smiles and satisfied, the baby resumes playing with mother.

In other words, babies take an intentional stance in order to share their inner experience with both their parents. This corresponds to E. BATES' definition of intentionality: "The sender is aware, in advance, of the effect that the signal will have on the receiver and will continue to act so as to obtain that expected goal" (BATES 1979, p36).

Another illustration of this intentional stance in sharing of affect is *affect signaling* (BRUNER 1998). It concerns negative affects. In this three-person interaction, the baby boy, frustrated by what is happening with his mother, addresses her an anger signal, and turns to his father with the same signal, presumably hoping the latter will give in. The father shrugs his shoulders with a smile, signaling empathy but unwillingness to interfere with mother. The baby has to understand that he has to negotiate himself with mother. Note that many other possibilities were open: the father could have interfered with the mother, or on the contrary ignored the infant's bid or blamed him for asking. We have observed a consider-

able range of responses in families, from the most differentiated to the most hostile or contemptful.

Perhaps the most striking process interlacing affect and cognition is *social referencing* (KLINNERT/CAMPOS/SORCE/EMDE/SVEJDA 1983). It has been extensively researched in triadic interactions between mother, infant and inanimate objects. We also observed it in triangular interactions in the LTP, where it evolves in three steps : 1) The infant is surprised by an ambiguous event and does not know which meaning to attribute to it. For instance, one parent has surprised the baby by doing something that was not in keeping with his/her style and the baby is startled. 2) The infant turns to the other parent with a quizzical expression, as if asking what is the matter, how should he understand this? 3) The other parent answers (it is a split-second event, therefore he or she does it intuitively and pretty much unaware). Suppose he or she gives the infant a reassuring look; it is then likely that the infant will consider this as a positive event and resume playing with the parent who surprised him. In other words, the baby and his parents are co-constructing a meaning via intersubjective sharing of affects.

Co-construction of meaning

As secondary intersubjectivity emerges, parents also modify their responses. Doted with what is called "intuitive parenting behaviors" (PAPOUSEK/PAPOUSEK 1987), they begin to spontaneously answer to their child's signals by means of "affect attunement" (STERN 1985); the latter process conveys that they share not only his behavior but also his mental state, as illustrated in the following example of triadic interactions. Baby and mother take turns at tapping on the baby's seat with a given tempo. Baby is extremely attentive and he works very hard at taking his turns. At some point, as time has come to introduce a new variation in order to sustain the baby's interest, mother switches from imitation, or the simple matching of the infant's act, to affect attunement: she moves her head in tempo instead of tapping the seat. This subtle operation keeps the temporal and intensity contour of the act invariant, while translating it into another non-verbal modality. According to D. STERN, it amounts to conducting a (pre)-symbolic operation, namely referring to an inferred internal feeling. The infant responds with delight, showing that he gets the message. He and his mother have shared an inner state and thus established twosome communion between them.

In triangular interactions, we observe parental affect attunement in response to an infant's social referencing about the other parent. This process has considerable implications for the understanding of transgenerational transmission of affective patterns in families. Indeed, the parent who is consulted by the infant is then in a position to redefine the entire network of relationships in the triangle. Optimally, he or she might attune to the infant's inner experience in an adjusted and differentiated way while at the same time supporting his or her co-parent's act. But he or she could also manipulate the network of relationships by resorting to forms of attunement that are selective and/or distorting, as in the following example observed in the LTP. Father and baby have fun playing together with a glass. But mother blames father for having introduced this glass (we ask the parents to try playing without objects). As her turn comes, mother gets the baby with holding the glass. She removes it from the baby's hands, reflecting aloud: your father has given you this glass, what am I going to do with it? Thus the affective climate abruptly switches from positive to dissonant: mother's tone of voice is blaming and father is laughing at mother's embarrassment. The baby turns to her father with a puzzled look (social referencing), gets a mitigated smile, looks at her right hand in which she had been holding the glass and finally reorients to mother. Mother states in a severe voice: "this is not how you are supposed to play with a glass". In other words, the tension between the parents has been detoured onto the child (VOGEL/BELL 1960). Following mother's blaming, the child pouts. It is on the underlying inferred feeling that the mother attunes: "yes I know you don't like it. This is life" she says, as she imitates her baby's pout and adds a vocal complaint and a head gesture conveying empathy for the child's frustration. In other words, the mother has distorted the meaning of the infant's feeling in order to help regulate the tension with her co-parent. As this mother tended to invariably attune to her baby's pout, to the detriment of other expressions, one may presume that the infant learned that intersubjective communion was restricted to this particular kind of negative experience.

For a complete demonstration of the handling of a nine-month old infant of the system of triangular interactions, it is necessary to examine his triangular bids in the four different configurations of the system, including the one where he is the third party, witnessing his parents' direct interaction. The results of our exploratory study indicated that nine month-old infants already use triangular bids in their inter-

action with their parents (see FIVAZ-DEPEURSINGE/CORBOZ-WARNERY/FRASCAROLO 1998). However, we observed large individual differences in the number of different configurations they handled, in the number of triangular bids they made; likewise, we observed differences in the proportion of positive (affect sharing) to negative bids (affect signaling plus social referencing) they made. These differences were related to differences in the "coparenting alliance". The coparenting alliance refers to the degree of coordination father and mother reached in working together to guide the infant in playing, to adjust to her momentary signals and to appropriately validating her triangular bids.

In summary, infants learn very early "how to be a suject among subjects" (NADEL/TREMBLAY 1999) and handling of triangular relationships calls for cognitive as well as affective operations : the infant constructs a system of representations of these relationships on the basis of her affective experience (BONNET 1999).

Primary intersubjectivity

With the earliest months of infancy, we are entering a less differentiated system of communication, unmediated by (pre)-symbolic processes. It is based on perception, i.e., the capacity that allows us to guide our actions or know our environment on the basis of sensory information (BONNET 1999, p175). There is however debate on how to interpret the data at hand. The extreme conservative perspective considers that this early communication is non-intentional, driven by context, mostly reactive, and merely reflexive around birth. However, accumulating evidence points to a more progressive interpretation : there would exist a primary form of intersubjectivity that is innate (TREVARTHEN 1984) and that makes social cognition the developmental matrix for physical cognition (BUTTERWORTH 1995; STERN 1985). Thus there would exist primary and secondary levels of intersubjectivity, of intentionality, of conscience, etc.

Early imitation: The debate on imitation is illustrative. Early imitation, from birth on, is now acknowledged as an inescapable fact, although the growing evidence has been repeatedly dismissed in the name of indifferentiation. PIAGET had defined the cognitive criterium for true imitation as the ability to match facial displays, because it requires to match a visible face to an invisible one. He had only observed it by the end of the first year and considered it as going hand in hand with secondary circular reactions (stage 4 of object permanence, looking for a hidden object; see for a discussion: MELTZOFF/MOORE 1992).

But as evidence is accumulating on neonatal and later facial imitation (e.g., tongue and lip protrusion, mouth opening), another theory is emerging which considers imitation as innate and based on multimodal perception (BUTTERWORTH 1998). There would be progressively more differentiated levels of imitation. Instead of being merely reflexive (a release preset motor packet in the way of a Moro reaction) and of being devoid of intrapsychic function (the cognitive interpretation), early imitation would serve a social cognition function: reenacting an act produced by another person would serve to know, to identify the person and verify this identity by forming a memory linking the person and the action, somewhat in the same way that manipulation serves to explore inanimate objects (MELTZOFF/MOORE 1995).

Recent studies by MELTZOFF and MOORE (1992) argue against the reflexive theory and in favor of primary intersubjectivity; they showed for instance that the young infant can delay imitation and that there is no drop in facial imitation at three months in concert with the drop out of other reflexive responses; they also showed that young infants imitate static facial postures as well as dynamic displays, and that they match their temporal as well as their spatial features—too sophisticated an operation to be a mere reflex.

Another line of analysis emphasizes the fact that researchers have been asking the infant triangular questions for a long time but without taking note of it. TREMBLAY-LEVEAU remarks that experimental designs typically require the infant to compare and select between two objects or persons (TREMBLAY-LEVEAU 1997). These are paired comparison or habituation/dishabituation paradigms for studying the infant's perceptual preferences or her abilities to discriminate between two elements.

Finally, other arguments in favor of primary intersubjectivity are that precursors of coordination of attention in triadic interactions have been demonstrated as early as two months of age (gaze following, SCAIFE/BRUNER 1975) and more clearly at four months (D'ENTREMONT/HAINS/MUIR 1997).

Early triangular interactions

In favor of primary intersubjectivity, E. TRONICK has argued that the very young infant in dialogue with mother refers to the interaction itself as she actively regulates it (TRONICK 1981). This process is

more distinct in triangular interactions, where it doesn't need to be infered, as the following data show.

A sample of 20 four-month old infants have been observed in a modified version of the LTP, the "LTP with still face" (DONZÉ 1998; FIVAZ-DEPEURSINGE/FRASCAROLO 1999). Starting with a three-together configuration, we ask the infant whether she can share her attention and affects with her two parents when they are both engaging in coordinated play with her. The answer is positive. Not only do infants distribute their attention more or less equally between their parents, but they also make rapid transitions (gazing at one parent and at the other within a 3 seconds time window), associated with positive or negative affective signals that prefigure affect sharing and affect signaling at nine months. Rapid transitions imply that the infant is able to keep the sequence of events in her working memory (MUELLER/LUCAS 1975).

Proceeding with a two-plus-one, we ask the baby whether, having experienced the three-together, she will go on construing the two-plus-one as triangular also. Results indicate that she does, although triangular bids are not as frequent in the two-plus-one configuration; they are most likely to occur when the infant is frustrated.

The next part includes an experimental manipulation, in which we ask the previously active parent to pose a still face. The still-face paradigm is a well researched situation. It was designed to test the infant's expectancies about what goes on in a normal dialogue with her mother (TRONICK/ALS/ADAMSON/WISE 1978). Infants are typically surprised by the absence of reaction on the mother's part and make active efforts to activate the mother. Then, faced with her continuing absence of reaction, they progressively give up contacting her to devote their efforts to self-regulating; nevertheless, they keep monitoring her from the side. These results show that the very young infant is sensitive to violations of interactive rules and that precursors of inferential capacities concerning human intentionality are present (NADEL/TREMBLAY 1999).

In the triangular paradigm, we also wanted to test the ability of the infant to recourse to the other parent in this ambiguous situation. It would constitute a precursor of social referencing. The answer is, infants tend to turn to the other parent with distress signals, but few display at this age the clearly puzzled expression that will characterize social referencing at nine months. However, pilot data show that such bids have notably increased by five months of age (DONZÉ 1998).

Finally, we asked the infant how she would reconciliate with her parents after the stress of the still-face; in particular, would she turn to the parent who has not been posing a still face as a mediator in repairing her interaction with the formerly still-faced parent? Some did it, others kept away for a while and slowly repaired the interaction, some exclusively interacted with the formerly third party parent.

In summary, all four-month old infants conduct triangular coordinations of attention and of affect with their parents, but few do it in all configurations. The findings of our study of 20 three-month old infants observed in the regular LTP concurs with these results. So do other results on non-family triads at the same age; they show in particular that infants this age follow the gaze of one experimenter towards another experimenter, check back at the first one and address them affective signals (TREMBLAY-LEVEAU 1998). In other words, the young infant is building an understanding of what is going on between two other subjects and themselves. These findings confirm the theory of a primary form of intersubjectivity.

Early interaction and representation

It is established that affective processes cannot be separated from cognitive ones in early infancy. D. STERN (1985) states it this way: "Affective experiences have their own invariant and variant features. Sorting them is a cognitive task concerning affective experience" (p42).

Schematically stated, D. STERN views the infant as constructing "schemes of being with " as they interact behaviorally with other persons. Infants don't need words or symbols to do it since knowledge constructed in this way is non-verbal and procedural rather than declarative. The process by which they form these schemes is by pattern recognition, i.e., looking for repeated events. When something repeats itself, they pick out the invariants. From invariants, they put together a prototype of what it is like to be in that particular situation. Whether they are with one or two people, things are the

Author's address

Elisabeth Fivaz-Depeursinge, Centre d'Etude de la Famille, Département Universitaire de Psychiatrie Adulte, Site de Cery, CH-1008 Prilly, Switzerland.
Email: elisabeth.fivaz@inst.hospvd.ch

same, thanks to peripheral as well as focal perception. For a baby, there is no problem with having two things going on at the same time. Therefore, there is no problem in forming schemes of being-in-a-threesome as well as schemes of being-in-a-twosome (see for a more detailed discussion: STERN 1995; FIVAZ-DEPEURSINGE/STERN/CORBOZ/BÜRGIN 1998).

Conclusions

Recent theories have stressed that emotion and cognition are necessarily linked in order to promote adaptation in human beings, for instance L. CIOMPI's theory of affect logic (CIOMPI 1982; WIMMER/CIOMPI 1996), R. FIVAZ's morphodynamics (FIVAZ 1989, 1996) and A. DAMASIO's neuro-science theory (1994). Present evidence in early developmental science indicates that this link already operates in the first few months of life: very young infants experience intersubjectivity where affects drive cognitions that in turn initiate further affects. Thus, infants manifest intentional stances and coordinate their attention and affects with two persons much earlier than previously thought. These findings confirm that affect and cognition are inextricably interlaced and thereby initiate and support a development consistent with the social and physical environments.

References

Bakeman, R./Adamson, L. B. (1984) Coordinating Attention to People and Objects in Mother–Infant and Peer–Infant Interaction. Child Development 55:1278–1289.

Barrett, J./Hinde, R. A. (1988) Triadic interactions: Mother-first-born-second-born. In: Hinde, R. A./Stevenson-Hinde, J. (eds) Relationships within families. Mutual influences. Clarendon Press: Oxford, pp. 181–190.

Bates, E. (1979) The Emergence of Symbols. Academic Press: New York.

Bonnet, C. (1999) Les trois étapes de la perception. In: Dortier, J.-F. (ed) Le cerveau et la pensée. Presses Universitaires de France: Paris, pp 175–180.

Bruner, J. S. (1978) From Communication to Language: A Psychological Perspective. In: Marlova, I. (ed) The Social Context of Language. John Wiley & Sons: New York, pp. 17–48.

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. 35–51.

Butterworth, G. (1998) Origins of joint visual attention in infancy. Commentary. In: Monographs of the Society for Research in Child Development 63 (4, No 9, 255):144–166.

Campos, J. (1994) The new Functionalism in Emotion. SRCD Newsletter 1:1–14.

Carpenter, M./Nagell, K./Tomasello, M. (1988) Social Cognition, Joint Attention and Communicative Competence from 9 to 15 Months of Age. Monographs of the Society for Research in Child Development 63 (4, No 9, 255):PAGE-NUMBERS MISSING.

Ciompi, L. (1982) Affektlogik: Über die Struktur der Psyche und ihre Entwicklung. Ein Beitrag zur Schizophrenieforschung. Klett-Cotta: Stuttgart.

Damasio, A. (1994) Descartes' Error. Avon Books: New York.

D'Entremont, B./Hains, S./Muir, D. (1997) A demonstration of gaze following in 3- to 6-month-olds. Infant Behavior and Development 20(4):569–572.

Donzé, F. (1998) Etude exploratoire des comportements triadiques du bébé de 3–5 mois. DEA en Psychologie, Université de Genève.

Fivaz-Depeursinge, E./Corboz-Warnery, A. (1999) The Primary Triangle. A Developmental Systems View of Mothers, Fathers and Infants. Basic Books: New York. Translation in German (2001) Das primäre Dreieck. Vater, Muter und Kind aus entwicklungstheoretisch-systemischer Sicht. Carl-Auer-Systeme Verlag: Heidelberg.

Fivaz-Depeursinge, E./Corboz-Warnery, A./Frascarolo, F. (1998) The Triadic Alliance between father, mother and infant, its relations to the infant's handling of triangular relationships. Paper presented at the International Symposium of the Society for Behavioral Development. Bern, Switzerland.

Fivaz-Depeursinge, E./Frascarolo, F. (1999) The Attentional and Affective Sharing Abilities of 3–4 Month-Old Infants with both Parents during Triadic Play. Paper presented at the Society for Research in Child Development, Albuquerque NM.

Fivaz-Depeursinge, E./Frascarolo, F./Corboz-Warnery, A. (1996) Assessing the Triadic Alliance between Father, Mother and Infant at Play. In: McHale, J. P./Cowan, P. A. (eds) Understanding how family-level dynamics affect children's development: Studies of two-parent families. Jossey-Bass: San Francisco, pp. 27–44.

Fivaz-Depeursinge, E./Stern, D./Corboz-Warnery, A./Bürgin, D. (1998) Wann und wie das familiale Dreieck entsteht: Vier Perspektiven affektiver Kommunikation. In: Welter-Enderlin, R./Hildenbrand, B. (eds) Gefühle und Systeme. Carl-Auer-Systeme: Heidelberg, pp. 119–154.

Fivaz, R. (1989) L'Ordre et la Volupté. Presses Polytechniques Romandes: Lausanne.

Fivaz, R. (1996) Ergodic Theory of Communication. Systems Research 13(2):127–144.

Kasari, C./Sigman, M./Mundy, P./Yirmya, N. (1990) Affective Sharing in the Context of Joint Attention. Interactions of Normal, Autistic, and Mentally Retarded Children. Journal of Autism and Developmental Disorders 20(1):87–100.

Klinnert, M. D./Campos, J. J./Sorce, J. F./Emde, R. N./Svejda, M. (1983) Emotions as behavior regulators: social referencing in infancy. In: Plutchik, R./Kellerman, H. (eds) Emotion. Theory, research and experience. Academic Press: New York, pp. 57–86.

Meltzoff, A. N./Moore, M. K. (1992) Early imitation within a functional framework: The importance of person identity, movement and development. Infant Behavior and Development 15: 479–505.

Meltzoff, A. N./Moore, M. K. (1995) A Theory of the Role of Imitation in the Emergence of Self. In: Rochat, P. (ed) The

Self in Infancy: Theory and Research. Elsevier: Amsterdam, pp. 73–93.

Mueller, E./Lucas, T. (1975) Developmental analysis of peer interaction among toddlers. In: Lewis, M./Rosenblum, L. (eds) Friendship and Peer Relations. Wiley: New York, pp. 223–257.

Nadel, J./Tremblay-Leveau, H. (1999) Early perception of social contingencies and interpersonal intentionality: Dyadic and triadic paradigms. In: Rochat, P. (ed) Early social cognition. Lawrence Erlbaum: Mahwah NJ, pp. 189-212.

Pacteau, C. (1999) La genèse des sens: Le foetus au nouveau-né. In: Dortier, J.-F. (ed) Le cerveau et la pensée. PUF: Paris, pp. 181–186.

Papousek, H./Papousek, M. (1987) Intuitive Parenting: A dialectic counterpart to the infant's integrative competence. In: Osofsky, J. D. (ed) Handbook of Infant Development, 2nd edition. Wiley: New York, pp. 669–720.

Papousek, H./Papousek, M./Koester, L. (1986) Sharing emotionality and sharing knowledge: A microanalytic approach to parent-infant communication. In: Izard, C./Read, P. (eds) Measuring emotions in infants and children. Cambridge University Press: Cambridge, pp. 93–123.

Parke, R. D./Power, T. G./Gottman, J. M. (1979) Conceptualizing and quantifying influence patterns in the family triad. In: Lamb, M. E./Suomi, S. J./Stephenson, G. R. (eds) Social interaction analysis: Methodological issues. University of Wisconsin Press: Madison, pp. 207–230.

Scaife, M./Bruner, J. (1975) The capacity for joint visual attention in the infant. Nature 253: 265–266.

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

Stern, D. N. (1995) The Motherhood Constellation. Basic Books: New York.

Tomasello, M./Call, J. (1997) Primate Cognition. Oxford University Press: Oxford.

Tremblay-Leveau, H. (1997) La triade: Une nouvelle matrice de développement? Rapport d'habilitation, Université de Rouen.

Tremblay-Leveau, H. (1998) Sharing attention with another on a third person in 3 and 6 month-old infants. Paper presented at the International Symposium of the Society for Behavioral Development. Bern, Switzerland.

Trevarthen, C. (1984) Emotions in infancy: Regulators of contact and relationships with persons. In: Scherer, K. R./Ekman, P. (eds) Approaches to emotion. Lawrence Erlbaum: Hillsdale NJ, pp. 129–157.

Trevarthen, C./Hubley, P. (1978) Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In: Lock, A. (ed) Action, gesture and symbol. The emergence of language. Academic Press: New York, pp. 183–229.

Tronick, E. (1981) Infant communicative intent: The infant's reference to social interaction. In: Stark, R. E. (ed) Language Behavior in Infancy and Early Childhood. Elsevier: New York.

Tronick, E./Als, H./Adamson, L./Wise, S./Brazelton, T. B. (1978) The infant's response to entrapment between contradictory messages in face-to-face interaction. Journal of the American Academy of Child Psychiatry 17: 1–13.

Wimmer, M./Ciompi, L. (1996) Evolutionary Aspects of Affective-Cognitive interactions in the light of Ciompi's concept of "Affect-Logic". Evolution and Cognition 2(1):37–57.

Shulamith Kreitler

Consciousness and States of Consciousness

An Evolutionary Perspective

Abstract

The goal of the paper is to present a new approach to the definition of consciousness and states of consciousness in terms of the meaning system and to outline its evolutionary background. Starting with the psychosemantic conception of cognition as a meaning-processed and meaning-processing system, the major constituents and functions of meaning are described in line with the theory of KREITLER *and* KREITLER*. In accordance with this theory it is suggested that states of consciousness are the product of meaning-prompted organizational transformations of cognition that affect the whole of the cognitive system and bring about changes also in other systems, mainly behavioral, emotional, personality and the self. Further elucidation of the conception is attained by tracing evolutionary developments of meaning and of the organizational transformations corresponding to states of consciousness.*

Key words

Meaning, consciousness, states of consciousness, cognition.

Introduction

The purpose of this paper is to present an analysis of consciousness (CON) and states of consciousness (SOC) from the evolutionary point of view, in an attempt to gain deeper understanding into their nature, properties and dynamics. The analysis is motivated by the conviction that far from being an appendage, artificially superimposed on the issue of interest, the evolutionary approach is an integral part of the conception without which no adequate understanding and presentation of an issue in the human sciences can be complete. In other words, since CON exists on the human level there can be little doubt that it represents a fundamental biological adaptation, whose nature and function will be better understood from the evolutionary point of view (BAARS 2001). However, due to scarcity of information about different aspects, the analysis will be at this stage preliminary, based on available data, conclusions projected from observations of ontogenetic development and hypothesized links.

Consciousness and States of Consciousness: Definitions and Assumptions

The analysis of the evolution of CON and SOC will be carried out in the framework of the cognitive approach to CON and SOC and will be guided by several basic assumptions. The first and major one is that CON and SOC are cognitive phenomena. Indeed, CON and SOC have often been discussed in relation to cognition. Thus, cognition is mostly presented as the object of CON, the substance or matter with which CON deals, namely, concepts, sensations, perceptions, moods, and dreams. Further, cognition (namely, language, high-level processing, etc.) is described as the antecedent, condition or cause for CON. And last but not least, cognition is assumed to be the function of CON, for example, performance of specific mental operations, control of mental states, focusing of attention, and so on (BAARS 1988). However, our assumption goes beyond these observations insofar as it expresses the view that CON and SOC are the products of specific properties and changes in the cognitive system itself.

The second assumption is that CON and SOC are closely related concepts, in fact variants of the same concept. According to the commonly accepted approach CON and SOC denote distinct phenomena. CON is considered as the basic dominant property which however is not always present because some functions or contents may not need CON or may be barred from CON for emotional reasons, in which case they are described as being preconscious or unconscious. Further, SOC are assumed to be products

of specific agents or phenomena external to CON and cognition, such as drugs or physiological phenomena that bring about the emergence of so-called altered SOC. In contrast, our second assumption is that there is only one basic property that characterizes the state of the cognitive system so that the cognitive system can be assumed to be always in a state of CON. States of CON include as one particular example what is commonly called CON, as well as all of the SOC identified so far as well as those that may be identified or that will emerge in the future. Hence all SOC may be considered on the same conceptual level, without any need to distinguish between CON and SOC or between CON or SOC and altered SOC.

The two cited assumptions form the basis for the elaboration of the evolutionary perspective. Therefore, we will present first, however briefly, the background for these assumptions which is the cognitive theory of CON that is rooted in the psychosemantic approach to cognition.

The Psychosemantic Approach to Cognition

Meaning: Definition and components

Our basic conception of cognition is psychosemantic. This approach indicates that meaning defines the essential contents and functioning of cognition. Meaning is a procedure for using cognitive contents for defining, expressing and communicating meanings for a variety of purposes, e.g., problem solving, comprehension, or communication. Meaning consists of meaning units, which include two components: 'the referent' which is the input, the stimulus, or the subject to which meaning is assigned, and 'the meaning value' which is the cognitive contents designed to express or communicate the meaning of the referent. The following are four examples of meaning units: 'table—serves for eating', 'bread—is on the table', 'milk—is produced by cows', 'bottle—is made of glass'. In these meaning units, 'table', 'bread', 'milk' and 'bottle' are the referents and 'serves for eating', 'is on the table' 'is produced by cows' and 'is made of glass' are the meaning values. Each meaning unit may be characterized in terms of meaning variables of the five following classes: meaning dimensions—which characterize the contents of the meaning values (e.g., locational qualities, material), types of relation—which characterize the immediacy of the relation between the referent and the meaning value (e.g., attributive, metaphoric–symbolic), forms of relation—which characterize the logical–formal properties of the relation between the referent and the meaning value (e.g., positive, conjunctive, partial), shifts of referent—which characterize the relations of the present referent to the initial input and previous referents (e.g., identical, partial, opposite), and forms of expression—which characterize the media of expression of the referent and/or the meaning value (e.g., verbal, graphic, motional). The meaning system consists of the whole set of the meaning variables.

Structure of the system of meaning

As noted, the system of meaning includes variables of five groups. Of the total of 90 meaning variables, 33.3% are meaning dimensions, 17.8% types of relation, 17.8% forms of relation, 14.4% shifts of referent and 16.7% forms of expression. Each of the five sets is complete in itself and independent of the other sets. Thus, characterizing a meaning unit involves using one variable from each set. Hence, when we have a group of meaning units characterized in terms of meaning variables and we count the frequencies of meaning variables used in characterizing these meaning units, we get in fact five independent groups of frequencies, namely, one for meaning dimensions, one for types of relation, one for forms of relation, one for shifts of referent, and one for forms of expression. Each of these five groups of frequencies amounts to the same total but consists of different meaning variables.

Each group of meaning variables has a unique structure. Meaning dimensions form a circumplex structure. Accordingly, they may be ordered in line with the similarity in their contents along the circumference of a circle, so that the more similar the two meaning dimensions are the closer they are placed to each other. The arrangement is based on varied data sources, including studies of multidimensional scaling, characteristics of participants' responses and testing of hypotheses based on contents similarity (KREITLER/KREITLER 1991). Further, the circular arrangement is defined in terms of two major axes recurring in most studies: an approximately vertical axis of abstractness–concreteness, anchored on the poles of contextual allocation and sensory qualities, and an approximately horizontal axis of action–emotion, anchored on the poles of action and feelings and emotions. In addition, each meaning dimension represents towards the center of the circle more general and global meaning values which become increasingly differentiated the

Meaning dimensions	Forms of relation
Dim. 1 Contextual Allocation Dim. 2 Range of Inclusion (2a: Sub-classes; 2b: Parts) Dim. 3 Function, Purpose & Role Dim. 4 Actions & Potentialities for Actions (4a: by referent; 4b: to, with referent) Dim. 5 Manner of Occurrence & Operation Dim. 6 Antecedents & Causes Dim. 7 Consequences & Results Dim. 8 Domain of Application (8a: as subject; 8b: as object) Dim. 9 Material Dim. 10 Structure Dim. 11 State & Possible changes in it Dim. 12 Weight & Mass Dim. 13 Size & Dimensions Dim. 14 Quantity & Mass Dim. 15 Locational Qualities Dim. 16 Temporal Qualities Dim. 17 Possessions (17a) & Belongingness (17b) Dim. 18 Development Dim. 19 Sensory Qualities (19a: of referent; 19b: by referent) Dim. 20 Feelings & Emotions (20a: evoked by referent; 20b: felt by referent) Dim. 21 Judgments & Evaluations (21a: about referent; 21b: by referent) Dim. 22 Cognitive Qualities (22a: about referent; 22b: by referent)	FR 1 Propositional (1a: Positive; 1b: Negative) FR 2 Partial (2a: Positive; 2b: Negative) FR 3 Universal (3a: Positive; 3b: Negative) FR 4 Conjunctive (4a: Positive; 4b: Negative) FR 5 Disjunctive (5a: Positive; 5b: Negative) FR 6 Normative (6a: Positive; 6b: Negative) FR 7 Questioning (7a: Positive; 7b: Negative) FR 8 Desired, wished (8a: Positive; 8b: Negative)
	Shifts in referent
	SR 1 Identical SR 2 Opposite SR 3 Partial SR 4 Modified by addition SR 5 Previous meaning value SR 6 Association SR 7 Unrelated SR 8 Verbal label SR 9 Grammatical variation SR 10 Previous meaning values combined SR 11 Superordinate SR 12 Synonym (12a: in original language; 12b: translated in another language; 12c: label in another medium) SR 13 Replacement by implicit meaning value
Types of relation	**Forms of expression**
TR 1 Attributive (1a: Qualities to substance; 1b: Actions to agent) TR 2 Comparative (2a: Similarity; 2b: Difference; 2c: Relationality; 2d: Complementariness) TR 3 Exemplifying-Illustrative (3a: Exemplifying instance; 3b: Exemplifying situation; 3c: Exemplifying scene) TR 4 Metaphoric-Symbolic (4a: Interpretation; 4b: Conventional metaphor; 4c: Original metaphor; 4d: Symbol) *Modes of meaning* Lexical mode (TR1+TR2) Personal mode (TR3+TR4)	FE 1 Verbal (1a: Actual enactment; 1b: Verbally described; 1c: Using available material) FE 2 Graphic (2a: Actual enactment; 2b: Verbally described; 2c: Using available material) FE 3 Motoric (3a: Actual enactment; 3b: Verbally described; 3c: Using available material) FE 4 Sounds & Tones (4a: Actual enactment; 4b: Verbally described; 4c: Using available material) FE 5 Denotative (5a: Actual enactment; 5b: Verbally described; 5c: Using available material)

Table 1: Major variables of the meaning system.

closer they lie to the external circumference of the circle (see Figure 1).

The types of relation also form a circumplex structure. Figure 2 shows the proximity placements on the circumference of the circle as well as the two major axes describing the arrangement. Here too the meaning values placed closer to the outer circumference are increasingly more differentiated (KREITLER/KREITLER 1985a).

The forms of relation form three sets of meaning variables. One set specifies the nature of the relation between the referent and the meaning value which may be existential–real (propositional), normative (required, necessary), questioned (does it exist?), or desired–wished for. A second set specifies the quantitative nature of the relation, which may be universal–absolute or partial. A third set specifies the relation when more than one meaning value is

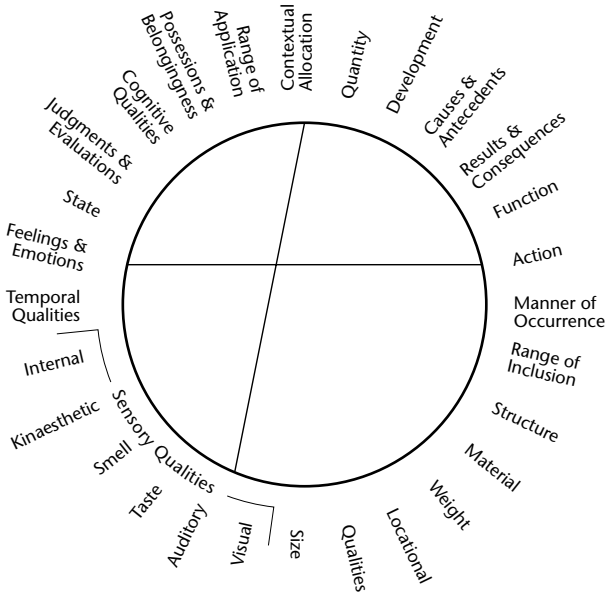

Figure 1: The Circumplex Model of Meaning Dimensions. The figure represents schematically the relations between the meaning dimensions in the system of meaning that seem likely on the basis of data available up to date. Some of the relations are still merely hypothesized. The locational position of the meaning dimensions represents their proximity. The closest relations are between adjoining variables, the furthest are between variables placed opposite each other on the circumference of the circle. The two intersecting lines represent factors identified in several studies. The variables at opposite poles represent meaning dimensions with positive and negative loadings on the factors, respectively.

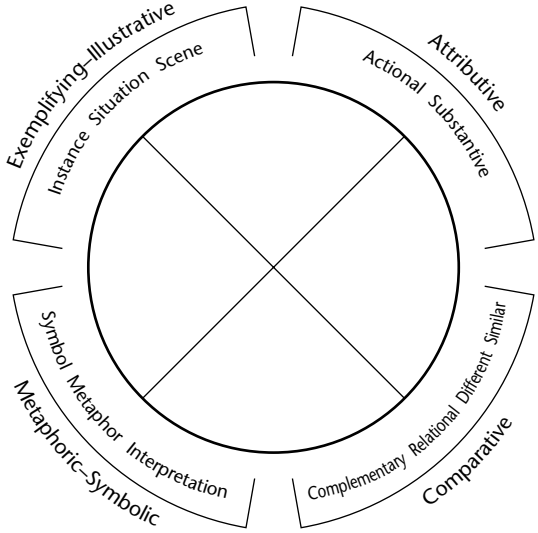

Figure 2: The Circumplex Model of Types of Relation. The figure represents in a schematic way the relations between the types of relation in the system of meaning that seem likely on the basis of available data up to date. The locational position of the types of relation represents their proximity. The closest relations are between adjoining variables, the furthest are between variables placed opposite each other on the circumference of the circle. The two intersecting lines represent factors identified in several studies. The variables at opposite poles represent types of relation with positive and negative loadings on the factors, respectively.

involved, so that it may be conjunctive or disjunctive. Each of the specified variables may be positive or negative.

The referent shifts form three classes of variables. One class includes the referent shifts to referents close to the original one (i.e., identical, partial, the original referent plus an addition, grammatical variation, synonym). A second class includes referent shifts to referents at medium distance from the original one (i.e., opposite, a previous meaning value, an association, a superordinate referent). A third class includes referent shifts to referents relatively far removed from the original one (i.e., unrelated referent, verbal label, a combination of several previous referents, an implicit unstated meaning value of the original referent).

The forms of expression include five sets which differ primarily in the mode of expression: verbal, graphic, movement, sounds, denotation. Each set includes variables which differ in whether the medium is used directly, verbally described or applied by means of materials.

Meaning: Static and dynamic properties

The description of the components of meaning indicates that it is a system, that it is complex, and that its elements are defined in terms of other elements of the system (namely, it is self-embedded and regressive). These three characteristics reflect the static or structural aspects of the system. They need to be complemented by three further properties that describe the dynamic aspects of meaning: it is a developing system both ontogenetically and phylogenetically; it is a selective system dependent in its structure and functioning on properties of the species, the individual and the input; and it is a dynamic system, whose special characteristics become manifest when it is activated for meaning assignment.

The connection between the static and dynamic properties of the meaning system takes place within the individual. Each individual disposes over a certain selected part of the meaning system which represents the specific tendencies of that individual to apply the meaning system in information process-

ing. Thus, each individual tends to use specific meaning variables with higher frequency and other meaning variables with medium or low frequency. The frequencies with which the individual tends to use each meaning variable are assessed by means of a test (The Meaning Test) and constitute the individual's meaning profile.

Meaning assignment

The meaning system functions primarily by providing the contents and processes enabling meaning assignment to inputs. Indeed, all the more elaborate functions of meaning depend on meaning assignment. Moreover, any act of human behavior in any domain requires prior identification of the input and sometimes also the input situation. Meaning assignment may be limited to one or several meaning values that enable straightforward limited identification of the input, for example, as a stimulus for a conditioned response or as the carrier of a particular label. But it also ranges all the way to more complex elaborations of meaning necessary for categorizing, classifying and evaluating material, as well as for problem solving and creative acts in different domains.

The contents and the extent of meaning assignment depend on several factors, major among which are the input, the context and the individual. The input may affect meaning assignment by means of its properties, such as familiarity or dangerousness and salient characteristics which attract attention to specific qualities or trigger particular meaning values (e.g., brilliant colors, loud tones, bizarre structure). The context affects meaning assignment by means of further inputs it presents and by contributing to particular definitions or interpretations of the input. Finally, the individual affects meaning assignment through the richness of his or her meaning system, the potential availability of meaning values and variables, as well as through personality traits and cognitive styles that may support meaning assignments of specific nature (e.g., emphasizing possessions, or internal stimuli) or of larger or more limited extent (e.g., cognitive needs for extensive meaning assignment) (KREITLER/KREITLER 1985b).

Meaning assignment is a gradually unfolding process, with several characteristic stages. A microgenetic study with visual inputs showed that the first phase was focused on primary input identification. It consisted of grasping specific meaning values in a specific order, referring to brightness, locational qualities and quantity, and ended in integrating the information in the form of a referent. The second phase was focused on elaborating and specifying the referent. It consisted of grasping specific meaning values dependent on the referent, namely, state and action if the referent was animate (animal, human), and size and structure if the referent was an inanimate object. These two phases are largely homogenous across individuals. The third phase was focused on a personal elaboration of the referent, amplifying the meaning of the referent by adding further referents and meaning values in line with individual tendencies (KREITLER/KREITLER 1984, 1986c). It is evident that not always all three phases take place. External and internal constraints (e.g., time pressure, danger, restricted meaning system) may limit the phases to one or two only.

In the above described study the participants were required explicitly only to identify the input. Thus, the meaning of the situation excluded from the very beginning the need for action. Under regular conditions meaning assignment appears to consist essentially in attributing or binding meaning values to a referent for the purpose of specifying the referent and the situation in which it is embedded in order to enable a decision whether the individual is involved in a form and to an extent that requires action on his/her part. If the primary input identification implies that a specified response of the conditioned or unconditioned kind is adequate for the situation, no further meaning assignment takes place. If meaning assignment implies that no conditioned or unconditioned response is or has been adequate, meaning assignment is elaborated further, focused on clarifying whether another kind of action is required. If that is the case, the further meaning elaboration is geared to specifying the kind of action (i.e., which particular emotion or cognitive act or behavior etc.). Again, the meaning elaboration of the intent for a specific action results in the specification of an action strategy or some plan for implementing the intent (KREITLER/KREITLER 1982).

Thus, the contents and extent of meaning assignment depend both on the situation and on the individual. When the situation is highly constrained, meaning assignment will remain focused on the demand characteristics of the situation, if the individual's meaning system enables responding in line with the requirements. When the situation is less constrained, the individual's meaning assignment will manifest to a greater extent his or her meaning assignment tendencies (KREITLER/KREITLER 1985b).

The effects of the situation are to be considered within the broader framework of context. The context includes both the external situation in which the input is embedded as well as the internal situation of the individual (e.g., mood, fatigue, concentration). Moreover, context includes not only the external and internal situations at the point in time in which the meaning assignment takes place but may also extend to previous states insofar they are stored in memory or affect the present state. For example, previously evoked emotions, or previously performed cognitive acts may predispose the individual to specific responses to a new input. Studies in which specific inputs (i.e., words) were presented for meaning assignment without context and then again in different contexts showed that there were different degrees of the effects of context on meaning assignment. The effects of context ranged from no change in the meaning of the input, to a contraction in the meaning, expansion in the meaning, and in the highest degree to a complete transformation of the input's meaning (as, for example, in metaphoric meaning). The effects depended on the properties of the context—its interestingness, conventionality and novelty (KREITLER/KREITLER 1993b).

Functions of meaning

The major and most essential function of meaning is input identification. This function is implemented by providing the contents and processes enabling meaning assignment to inputs. As noted, input identification ranges all the way from limited identification in terms of a stimulus for a particular action to highly complex meaning elaborations necessary for acts involving cognitive, emotional, physiological and behavioral components. Input identification is the prior condition for any further action on any level. If a personality trait is to be enacted or an emotion is to be evoked depends primarily on how the input has been identified. For example, an input that has been identified as related to success or failure is likely to enable enactment of achievement-related traits or emotions such as pride in achievements, or fear of failure.

A further function of the meaning system is to provide the cognitive contents and processes necessary for carrying out different cognitive acts. Studies showed that each meaning variable represents a specific set of contents and processes. For example, the meaning dimension Locational Qualities represents the set of contents denoting location (e.g., special, geographic) and the processes involved in dealing cognitively with locations (e.g., identifying, specifying, recalling, transforming locations). Further studies showed that each type of cognitive act corresponds to a specific pattern of meaning variables that provide a description of the contents and processes involved in its enactment. For example, meaning variables involved in planning include structure, temporal qualities, and causes and antecedents (KREITLER/KREITLER 1986a, 1987b,c). If the individual's meaning profile includes a sufficient proportion of the meaning variables included in the pattern corresponding to the particular cognitive act, that individual will be able to perform well the particular cognitive act (KREITLER/KREITLER 1986d, 1989, 1990a, 1994).

A third function of the meaning system is manifested in the domain of personality. A body of research showed that each of over 200 personality traits corresponds to a specific pattern of meaning variables. Again, as in the case of cognitive acts, the pattern of meaning variables may be considered as providing a description of the contents and processes involved in the enactment of the specific trait. For example, the meaning variables in the pattern corresponding to extraversion include high salience of the meaning dimensions of action, sensory qualities, temporal qualities and belongingness of objects, as well as low salience of the meaning dimensions of internal sensations and cognitive qualities (KREITLER/KREITLER 1990b, 1997). If the individual's meaning profile contains a sufficient proportion of the meaning variables included in the pattern corresponding to the particular personality trait, it is highly likely that the individual scores high on that personality trait.

The same holds in regard to further tendencies in the domain of personality, such as personality dispositions, defense mechanisms (KREITLER/KREITLER 1993a) and even the self. An individual's self concept was shown to consist of contents representing the major meaning variables in the individual's meaning profile (KREITLER/KREITLER 1987d; KREITLER in press-a).

The fourth function of the meaning system is in regard to emotions. Here too studies showed that particular patterns of meaning variables correspond to specific emotions, such as anxiety or fear. Further, changing experimentally specific components of the pattern corresponding to anxiety decreased the level of anxiety and its effects (KREITLER/KREITLER 1985a, 1987a).

In sum, the four functions of the meaning system that have been identified indicate that the meaning

system provides the understructure—that is, the raw materials in terms of contents and processes—for input identification, cognitive functioning, personality tendencies and emotions. All four functions depend on meaning assignment and reflect the central role of meaning for and within cognition. This has given rise to the psychosemantic conceptualization of cognition as a meaning-processing and meaning-processed system.

Meaning-mediated changes in cognition

Since cognition is a functional system, changes may be expected to take place in it most of the time. The changes are mediated through meaning and may therefore best be described in terms of meaning. All changes consist in changes of meaning values. However, they differ in four major respects. First, the changes vary in their *nature*. Some changes consist in shifting from one meaning value to another within the same meaning variable (e.g., in the meaning dimension Locational Qualities from 'in Austria' to 'in the USA'). Other changes consist in shifting from a meaning value in one meaning variable to one in another meaning variable within the same set (e.g., shifting from the meaning dimension Locational Qualities to the meaning dimension Temporal Qualities). In regard to changes of meaning variables it is to be emphasized that some meaning variables are related more closely to specific other meaning variables so that when one becomes salient the related ones also become more salient than they would otherwise (e.g., graphic form of expression and the meaning dimension structure, verbal form of expression and the meaning dimension Temporal Qualities). Second, the changes differ in *complexity*. Some changes consist in changing only one meaning value whereas others involve changes in several meaning values or even several meaning variables. Third, the changes may vary in *size*. A change is small if it consists in shifting between meaning variables that are close to each other, such as meaning dimensions located in adjoining positions in the circumplex model of dimensions. The change becomes larger if the number of intervening meaning variables characterizing the two meaning values increases (e.g., from comparative: similar to exemplifying–illustrative: instance, see Figure 2). Finally, the changes differ in their *originating cause*.

There are two major causes for the changes. One cause originates in the cognitive system and has to do with performing some task which is externally presented (e.g., solving a problem or making a decision) or arises from the needs of the cognitive system itself (e.g., organizing material). Performing the task brings about changes in meaning, such as focusing on meaning variables that promote the solution and de-emphasizing the irrelevant or dysfunctional ones. These so-called cognition-prompted changes are characteristically limited in extent and duration.

Another cause of changes originates in the meaning system and has to do with structural reorganizations of the system, spurred for example by recurrence of similar tasks, or confrontation with recurring difficulties or accumulation of material that is incompatible with the existing organization. As in the case of cognition-prompted changes, the changes consist in placing different meaning values or variables in a focal position while neutralizing others or relegating them to a secondary position, but the changes mostly involve a greater number of meaning values or components, and are designed to serve a broader range of tasks than just the specific task at hand. Hence, they are not dissolved after a particular task but are mostly stored and can be retrieved under appropriate circumstances. These meaning-prompted changes may best be called organizational transformations (see Table 2 for examples of organizational transformations).

Organizational transformations may arise in response to the needs of the meaning system itself, but may be evoked also in response to the cognitive system as well as other systems, through the intermediation of cognition, primarily the emotional, personality, the self and even physiological.

Organizational transformations bring about changes in the nature, salience and interconnectedness of contents and of cognitive processes as long as the transformations are activated. Thus, some contents and processes may become highly salient (e.g., Feelings and Emotions in the 'Emotional Approach') whereas other contents and processes may become hardly available or outright blocked (e.g., sensory information from the inner body in the 'Abstract Approach') (see Table 2). Changes of this kind affect most clearly ongoing cognitive functioning, for example, facilitating or hampering the solving of problems of specific types. But in addition they bring about also changes in other systems dependent on cognition, primarily personality traits and dispositions, emotions, or the self concept.

The organizational transformation	Focal meaning variables	De-emphasized meaning variables
Abstract approach	Dim1 Contextual Allocation; Dim2a Range of Inclusion (subclasses); Dim6 Antecedents & Causes; Dim7 Consequences & Results; Dim21 Judgments & Evaluations. TR1 Attributive; TR2 Comparative. FR1 Propositional. SR1 Identical; SR3 Partial; SR4 Modified by addition; SR9 Grammatical variation; SR11 Superordinate. FE1 Verbal.	Dim19 Sensory Qualities; Dim20 Feelings & Emotions. FR8 Desired, wished. SR7 Unrelated; SR8 Verbal Label; SR13 Replacement by implicit meaning value. FE2 Graphic; FE3 Motoric; FE4 Sounds & Tones; FE5 Denotative.
Concrete Thinking	Dim2b Range of Inclusion (parts); Dim3 Function, Purpose & Role; Dim4 Actions & Potentialities; Dim9 Material; Dim12 Weight & Mass; Dim13 Size & Dimensions; Dim15 Locational Qualities; Dim19a Sensory Qualities (of referent). TR3 Exemplifying Illustrative. FR1 Propositional; FR3 Universal. SR1 Identical; SR3 Partial; SR6 Association. FE2 Graphic; FE3 Motoric; FE4 Sounds & Tones;FE5 Denotative.	Dim1 Contextual Allocation; Dim2a Range of Inclusion (subclasses); Dim6 Antecedents & Causes;Dim7 Consequences & Results; Dim21 Judgments & Evaluations; Dim22 Cognitive Qualities. TR1 Attributive. FR7 Questioning. SR11 Superordinate. FE1 Verbal.
Emotional approach	Dim11 State & Possible change in it; Dim19b Sensory Qualities (by referent);Dim20 Feelings & Emotions. TR3 Exemplifying Illustrative. FR6 Normative;FR7 Questioning; FR8 Desired, wished. SR4 Modified by Addition; SR6 Association; SR7 Unrelated; SR13 Replacement by implicit meaning value.	Dim1 Contextual Allocation; Dim4 Actions & Potentialities. TR1 Attributive. FR1 Propositional. SR1 Identical; SR5 Previous Meaning Value; SR8 Verbal Label; SR12 Synonym.
Personal Mode*	Dim11 State & Possible change in it; Dim19b Sensory Qualities; Dim20 Feelings & Emotions; Dim21 Judgments & Evaluations; Dim22 Cognitive Qualities. TR3 Exemplifying Illustrative; TR4 Metaphoric–Symbolic. FR3 Universal; FR4 Conjunctive; FR6 Normative; FR8 Desired, wished. SR2 Opposite; SR3 Partial; SR4 Modified by addition; SR6 Association;SR7 Unrelated; SR10 Previous meaning values combined; SR13 Replacement by implicit meaning value. FE2 Graphic; FE3 Motoric; FE4 Sounds & Tones;FE5 Denotative.	Dim1 Contextual Allocation; Dim2 Range of Inclusion; Dim3 Function, Purpose & Role; Dim6 Antecedents & Causes; Dim7 Consequences & Results; Dim8 Domain of Application; Dim14 Quantity & Mass. TR1 Attributive; TR2 Comparative. FR1 Propositional;FR2 Partial; FR5 Disjunctive. SR1 Identical; SR11 Superordinate. FE1 Verbal.

Table 2: Examples of organizational transformations in the meaning system. Dim = Meaning Dimension, TR = Type of Relation, FR = Form of Relation, SR = Shift in Referent, FE = Form of Expression.
*) The Interpersonal Mode consists of the mirror-image of the Personal Mode, so that all the Focal meaning variables of the Personal Mode become relegated to the De-emphasizing position and vice versa. *Note:* In the case of meaning variables with subtypes (see Table 1), when no subtypes are specified, all subtypes are indicated.

A Cognitive Theory of Consciousness

Our suggestion is that a SOC is a product of a meaning-prompted organizational transformation that affects the whole of the cognitive system. One major characteristic of the definition is that the changes are meaning-prompted. This indicates that though they occur in the cognitive system they are due to and reflect an organizational transformation in the meaning system. Even when the changes are induced by conditions external to cognition and meaning, such as drugs, physiological states or a hypnotic technique, the changes that form the basis for a SOC are those that occur in cognition. The other important characteristic of the definition is that the changes affect the whole of cognition. Hence, SOC may be considered as reflecting a gestalt quality of the totality of the cognitive system.

Notably, the meaning-prompted changes in cognition affect the structure and functioning of the cognitive system but do not replace its intrinsically-motivated (i.e., cognitively-prompted) dynamics and functioning. Hence, the SOC affects the cognitive contents and processes that will be activated but not the precise tasks and their results. Accordingly, a SOC may be compared to mode in music that affects the kind of tones and scales that are being used but does not determine which music will be composed. Further, since the cognitive system affects further systems in the individual (e.g., emotions, personality, self), a SOC is reflected in changes in the total person (KREITLER 1999, 2001a).

A most dramatic demonstration of the suggested approach to CON is provided in a series of studies based on producing in individuals an organizational transformation of the meaning system for limited time duration and checking the manifestations thereof in different domains. The organizational transformation consisted in inducing a dominance of the personal mode of meaning or alternately of the interpersonal (lexical) mode of meaning. The induction procedure consisted in guiding the participant to respond in terms of the relevant types of relation (viz. the exemplifying–illustrative and metaphoric–symbolic for the 'personal mode') to a set of specific referent chosen as particularly evocative for the desired types of relation (e.g., mother or loneliness for the 'personal mode'). Under the impact of these inductions different tasks were administered. The results showed, for example, that under the sway of the personal mode induction normal participants produced a great number of so-called pathological responses in the Rorschach, their reality-testing declined, their creativity (as assessed by scores on fluency, flexibility, elaboration, and originality) increased, their gestalt perception was enhanced, their reaction time to discrimination of stimuli increased, their recall of faces improved, etc. Under the sway of the interpersonal (lexical) mode, solution of logical problems was improved, associations became more restricted, and control of emotions improved (KREITLER 1965; KREITLER 1999; KREITLER/KREITLER/WANOUNOU 1987–88; LAHAV 1982).

In principle, any organizational transformation in the meaning system may be considered as generating a SOC. Thus, there are an infinite number of possible SOCs. However, they vary in their impact. Some organizational transformations exert a dramatic, even shocking impact, whereas others are hardly noticeable and are experienced as mere fluctuations in the prevailing SOC.

The variations in impact are due to a number of factors. The first factor is the extent of the underlying changes. The extent is mainly but not exclusively a function of the number of meaning variables placed in focal position and the number of those weakened or blocked. Secondly, the nature of the changed meaning variables. Some meaning variables are of such focal importance in the functioning of the individual that their change is immediately noticeable whereas a change in a relatively weak or infrequently used meaning variable may be hardly noticeable. Thirdly, the duration of the change. Sometimes but not always the longer duration of the SOC contributes to enhancing its impact, though some SOCs that last only fractions of seconds may be noticed and even treasured for a lifetime. Fourthly, the extent of the changes in other systems of the individual. The more intense these changes are, especially the emotional ones, the stronger the impact. A fifth factor is the cultural interpretation attributed to some SOCs. SOCs sanctioned by the culture to which the individual belongs exert a stronger impact than they would otherwise (FABER 1981). For example, the differences between the four stages of SOCs characteristic of Zen meditation ('koan', 'sunmay', 'makyo' and 'satori') are probably hardly noticeable for an untrained member of Western culture (JOHNSTON 1971). On the other hand, it is likely that a SOC degraded or tabooed by a particular culture may also gain in impact precisely because of its relative unfamiliarity. Finally, a sixth factor is the difference between the particular SOC and the SOC habitual for the individual. The larger the difference, the

larger the impact which that SOC may be expected to exert.

Each SOC is in fact a discrete entity and is experienced as such. However, there may be different degrees of similarity between them, based on the kind and extent of the changes generated by the underlying organizational transformation. This may be the basis for groupings of SOCs into sets, such as the so-called states of consciousness in Western culture or the states of hypoarousal and tranquility in Eastern cultures (FISCHER 1978). The similarity may also determine the degree to which one may recall experiences of one SOC when operating under the sway of another SOC.

Our definition of SOC indicates that the property of being conscious is not limited to a particular SOC, say the so-called ordinary SOC (i.e., the SOC habitual of the Western normal educated adult of the 21st century when he or she happen to be awake and not drunk or otherwise intoxicated). Rather, consciousness is a property that denotes a specific degree of availability or readiness for evocation that concerns particular contents and varies from one SOC to another. Accordingly, in each SOC there are contents relegated to unconsciousness, specific to each Soc. Thus, in the ordinary SOC the unconscious contents are those that contradict the superego of the self-concept; in the hypnotic SOC the unconscious contents are those that are proclaimed to be unconscious by the order of the hypnotist or the individual (viz. self-hypnosis); in the daydream-dependent SOC the unconscious contents are those that do not correspond to the role playing assumed by the daydreamer (SINGER/ANTROBUS 1972); in the meditation-induced SOC the unconscious contents are defined by the technique that may dictate blocking perceptions of the external world (SINGER 1970).

Evolutionary Perspective on the Cognitive Approach to CON

There are several lines of evolution that need to be considered in order to shed light on the nature and function of SOC. We will outline development along two lines that seem to have major evolutionary contributions to the phenomenon at hand. These deal with meaning and with the organizational transformations in meaning. Meaning provides the raw materials underlying the phenomenon, and the organizational transformations provide the processes and structures that constitute the SOC.

Evolutionary perspective on the development of meaning

The structure and contents of meaning provide the basic materials for the operation of cognition that constitutes the framework for the emergence and functioning of SOC. It is evident that a primitive and limited meaning system cannot provide the underpinnings for organizational transformations required for SOC. Further, the more complex the meaning system is the richer the variety of SOC whose emergence and functioning it could enable. Hence it is of importance to trace the main evolutionary trends of meaning.

a. Development of meaning assignment. The basic components of meaning are meaning values. Meaning values are essentially cognitive contents, such as 'red', 'above', 'big', 'scary', 'eatable'. Notably, meaning values may be produced verbally but also in all nonverbal forms, including actions. The primary step towards the emergence of meaning occurs when cognitive contents are used for indicating meaning. This occurs when it is related or attributed to a stimulus which thus gets its meaning. Hence, at the stage when behavior is regulated primarily by internal homeostasis, there is no discriminated stimulus and hence there can also be no meaning (WIMMER 1995). The primary context in which meaning is related to a stimulus takes place is in producing a response to a stimulus, whereby the response may be as basic as a reflex or tropism. The response is in fact the meaning value, so that together with the stimulus—which in this context acts as referent—it gives rise to a primary meaning unit. This kind of meaning production was named action meaning (KREITLER/KREITLER 1982). It could be argued that since responses of the kind of reflexes or tropisms are inborn so that there is no choice or alternative in their evocation, the production of such responses does not count as meaning. The counter argument would be that factors such as freedom of choice, prior learning of the relation of response and stimulus, or intent are not necessary conditions for the production or use of meaning.

A further development in the meaning unit is attained when the response is not evoked automatically by the stimulus on the basis of inborn bonds but is a learned response, adapted to the stimulus as reflected also in the feedback-embedded evaluation of the response (LORENZ 1981). This applies to learned responses of the conditioned type. The conditioned response to a discriminated stimulus dem-

onstrates the development from the meaning value to the meaning unit.

The subsequent developmental stages of meaning assignment occur when several meaning values are attributed to one stimulus, for example in learning, when the stimulus evokes a hierarchy of responses. Meaning assignment changes then from act to process. One consequence of this change is that the stimulus in its role as referent undergoes changes resulting in greater elaboration and complexity (KREITLER/KREITLER 1986b). In later developmental stages, the sets or schemata of meaning values attached to specific referents may turn from rigid responses to options. This occurs as part of the extension of the number and nature of meaning values that may be evoked in regard to a referent. The meaning values become personalized and represent contents selectively bound to the referent in view of situational, personal and cultural constraints (KREITLER/KREITLER 1985b, 1988). This process has been called meaning generation (KREITLER/KREITLER 1982). Notably, at later developmental stages some of the meaning values, reflecting the lexical, culturally sanctioned part of the referent's meaning, again become attached to the referent and are evoked automatically. They may best be called label meaning and they constitute the core of the interpersonally-shared meaning which enables interpersonal communication.

In sum, this section has shed light on the development from cognitive contents to meaning value and from meaning value to meaning unit.

b. Development of number of components. As noted, the basic components of meaning are cognitive contents serving as meaning values. One major developmental trend consists in the enrichment of the number of potential meaning values. In fact, any cognitive contents may serve as meaning value after the function of meaning assignment has been acquired. Cognitive development at large consists in acquiring an increasing number of cognitive contents that could also serve as meaning values. The major processes enabling development of meaning values are differentiation and dissociation.

Differentiation consists in separating specific contents out of a complex of contents, for example, separating meaning values denoting function out of a complex denoting indiscriminately all actions. For example, differentiating between meaning values denoting function, such as 'shows the time' (for Clock), or 'defends my territory' (for Fence) and meaning values denoting actions, such as 'runs fast' (for Hare) or 'shouts' (for Boy next door). Similarly, meaning values of sensory qualities undergo differentiation into different subclasses and categories of sensations, both internal and external; feelings and emotions undergo differentiation into positive and negative emotions, as well as into moods and feelings.

Dissociation consists in liberating an item of cognitive contents from correlates or bonds associated with it rigidly. For example, specific meaning values are at first associated with specific forms of expression, for example in bees location is indicated by a specific movement. Bees are not free to communicate the meaning values of location by any other means. Later developmental stages enable dissociation of cognitive contents from specific meaning variables. This dissociation contributes to an increase in the number of meaning values.

Notably, both differentiation and dissociation are processes based on meaning variables: differentiation on the types of relation similarity and difference as well as the meaning dimension range of inclusion (subclasses), and dissociation on the negation of the form of relation conjunction. This observation demonstrates that the meaning system itself contributes to the development of its own components.

Meaning values constitute the raw materials of meaning. The development of the *system* of meaning requires organizing meaning values in terms of meaning variables, even if only implicitly (REBER 1993). This step is based on abstracting and categorizing. A meaning variable is a powerful generator of meaning values. Establishing a meaning variable enables producing many more meaning values of the specific kind represented by that meaning variable. It also provides better storing and retrieval possibilities, as well as operating with meaning values of the represented kind.

Studies of ontogenetic development show that in the first months following birth the human baby disposes over meaning values representing a broad range of what in later developmental stages would earn the name of meaning dimensions, e.g., Antecedents and Causes, Consequences and Results, Locational Qualities, Domain of Application. Hence, later development brings about the emergence of the different meaning dimensions as meaning variables and an increase in the number of meaning values in each of these meaning dimensions (KREITLER/KREITLER 1987c). With the increase in the number of meaning values of a particular meaning variable that variable gains in potency and operational potential in different domains.

Thus, human babies seem to have a fairly good start for their pilgrimage into the sphere of meaning. Of the 22 meaning dimensions, at least 15 are represented through meaning values observable in the first 6 months of life. In the human sphere, intelligence is correlated positively with an increased number of meaning variables (KREITLER, in press b). As may be expected, different species of animals seem to dispose over fewer meaning variables, though due to methodological problems it may be difficult to determine precisely how many meaning variables are actually represented. Phylogenetic development indicates, however, a progressive increase in the number of meaning variables representing the meaning values observable in different species of animals. Since this theme deserves a separate handling, only some examples will be cited. Thus, the number of meaning dimensions identified in bees seems to be 7–8 (VON FRISCH 1967), in ants 6–7 (WILSON 1971) and in chimpanzees as high as 13 (DE WAAL 1986).

The development of the meaning system consists in the emergence and stabilization of meaning variables of different types. It is likely that meaning variables of all five sets (Table 1) are acquired conjointedly. The development of a meaning variable consists in defining both the meaning values it represents or includes as well as those it does *not* include. Hence, by differentiation further meaning variables are formed.

As may be expected, there is a large variation in the meaning values of each meaning variable. Formation of a meaning variable entails over time also the subcategorizing of the different meaning values in the variable as well as their characterization in terms of a variety of criteria (e.g., intensity, durability).

An important developmental step consists in clarifying the relations between the meaning variables. This leads to two major consequences. One is the formation of sets of meaning variables. Meaning dimensions become organized into a set, as well as types of relation, and so on. The organization into sets is based on similarity of function between the meaning variables in the set. Thus, meaning dimensions characterize the contents of the meaning value, whereas types of relation characterize the immediacy of the bond between the referent and the meaning value. The second consequence is the emergence of structure within each set. This is based on defining the relations among the different meaning variables in the set. Defining inter-variable relations enables establishing degrees of proximity. The result is the emergence of a structure such as the circumplex that has been observed for meaning dimensions and types of relation (Figures 1 and 2). Proximity relations of this kind regulate the operation of the meaning variables within the set.

It is likely that the sequence in which the five sets underwent internal organization and structurization starts with meaning dimensions, followed by forms of relation because the latter represent relations of basic importance for communication (e.g., positive vs. negative, propositional vs. normative). Forms of expression may have come next perhaps because of the frequency of their use. Types of relation came probably later in the sequence because of the subtlety of the relations they represent. We assume that shifts of referent came last because they are based on prior development of meaning assignment into sequences of meaning units focused on one referent.

The organization and structurization of the sets of meaning variables enable a greater freedom of interrelating–conjoint, dissociated as well as selectively differentiated application of meaning variables within and across sets. This provides for greater cognitive power of the meaning variables for coping with diverse cognitive tasks of varying complexity and difficulty. But it also makes possible a further development of the meaning system. At higher developmental levels, the meaning system is a self-embedded system, that is, each of its part can serve as a focal point around which all the rest of the system is organized. The structure is self-unfolding. For example, when the meaning dimension Feelings and Emotions turns into a focus, then all the other meaning dimensions function as subdimensions representing subclasses, locational qualities, temporal qualities, sensory qualities etc. of Feelings and Emotions. The same holds in regard to the other sets of meaning variables that may subserve Feelings and Emotions. In this manner the focal meaning dimension undergoes enrichment in contents and structure. Hence, the self-embedded property of the meaning system turns the meaning system into a potent generator of further development of the meaning system.

In sum, this section has shed light on the development from meaning value to meaning variable and from meaning variable to meaning system.

c. Development of functional potential. Meaning is a dynamic system whose characteristic properties become evident when it is put into operation. The primary function of meaning is meaning as-

signment, initially manifested in regard to input identification. This basically limited act is cognitive but, as noted, its results serve the primary manifestations of behavior insofar as unconditioned and conditioned responses may be considered as manifestations of meaning assignment (viz. meaning action). The development of meaning assignment into a more elaborate process (viz. meaning generation) enables deeper involvement of meaning assignment in the sphere of behavior. When a greater number of meaning values, representing also personal meanings and those in deeper layers of personality, are drawn into operation, meaning assignment starts playing an increasingly important role in regard to resolutions about whether action will be undertaken or not, and if yes—which action and how it would be implemented (KREITLER 2001b).

Concomitantly the deepened meaning assignment to inputs extends also the role of meaning in the cognitive domain. The extended meaning assignment to referents in the cognitive sphere increases the involvement of meaning in problem solving (for example, by enabling the construction of the problem space), in decision making (for example, by enabling the specification of the alternative options and the elaboration o their meanings) and other cognitive acts. Progression of this involvement leads to the generation and increased involvement in cognitive functioning of patterns of meaning variables. Some of these patterns are more tightly structured sequentially and correspond to cognitive schemata, whereas others are less tightly organized and correspond to personally preferred dispositions resembling cognitive styles.

Cognitive styles could be considered as personality traits whose domain of operation is limited to the cognitive domain. They serve as paradigms for the development of patterns of meaning variables that provide the understructure for personality traits, other personality dispositions, the self and emotions. The selection of meaning variables, the organization and characteristics of the pattern as a whole kind differ for the different domains (e.g., emotions, physiology, personality traits). The reasons for the development of these patterns seem to be the tendencies for increased stabilization and regulation of action in the different domains. WIMMER (1995) showed how the involvement of symbolic information-processing mechanisms has increased the number, stability and regulation of emotions. The same functions have been fulfilled in the field of personality by the patterns of meaning variables corresponding to traits (KREITLER/KREITLER 1990b). Be it as it may, the increased involvement of meaning in different domains represents the functional extension of meaning into an expanding number of systems within the individual and possibly beyond.

In sum, this section has shed light on the development from static meaning values to dynamic meaning values and from dynamic meaning values to patterns of meaning variables affecting cognition and further systems in the individual.

Evolutionary perspective on the development of organizational transformations

As may be recalled, SOCs were defined as the product of meaning-prompted organizational transformations in the cognitive system. Such transformations represent elaborate changes in cognition and presuppose earlier developmental stages. An attempt at evolutionary reconstruction leads to the assumption that in the initial stage the cognitive system underwent only cognition-prompted changes due to cognitive tasks imposed externally or internally. Cognition-prompted changes presuppose a cognitive system at a minimum level of development, which is capable of at least simple cognitive acts and is supported by an at least rudimentary meaning system. There is little doubt that these conditions exist at the level of the lower mammals and possibly also of the higher insects.

At present it is only possible to hypothesize how meaning-prompted organizational transformations came about. It is likely that organizational transformations could have arisen as an extension of the state produced by a prolonged or a difficult cognitive task or in response to recurrent cognitive tasks of a certain type that produced similar changes in the cognitive system. It was useful and adaptive to store such changes for further use. Cognition-prompted changes could have been brought about also by non-cognitive recurrent stimuli, such as emotions, lack of sleep or food, intoxication or sickness.

Storing the changes could have been the initial manifestations of meaning-prompted transformations. The fact that the changes were stored and could be retrieved from memory made them gradually independent of the original task or situation that brought them about. Thus they underwent generalization in regard to usages—from one task to similar tasks and later to apparently non-similar tasks too. Their recurrent evocation could have led to structural modifications that produced a tighter and more stable organization.

Why were such organizational transformations stored and treasured? There seem to be two likely reasons. One is that specific cognitive functions were facilitated when particular organizational transformations were in effect. The other related reason is that specific cognitive, emotional, and personal experiences were more likely to happen when particular organizational transformations were dominant. These may have included on the human level mystical, revelatory (enlightment), and psychotic experiences or what appeared to be parapsychological phenomena. Hence, organizational transformations may have been promoted both in order to create optimal context conditions for particular cognitive acts as well as in order to gain informations and undergo experiences that otherwise would be difficult to come by (BENTALL 2000; WULFF 2000).

However it may be, the impact of organizational transformations on cognitive functionin and other systems is the culmination of a process that has started with the emergence of the organism's internal state as a determinant of action (WIMMER 1995). This followed a previous stage in which actions were automatically determined by impinging stimuli (LORENZ 1981). Intenal state enabled a loosening of the rigid bond between a stimulus and a particular response. One important consequence thereof was that the evocation of a particular response could depend on more than one stimulus, perhaps on two stimuli—first, a stimulus and another stimulus that has preceded it, then the stimulus and a collateral or associated stimulus in temporal or locational proximity. This is one possibility how context came to play a role in the evocation of a response and later also in the meaning assigned to the input. The introduction of context created the possibility for further processes, such as emotions to join and amplify the matrix of factors affecting the response. Thus, SOC could be considered as evolutionary extension of inner state.

Several evolutionary strands may be noted in regard to SOC. One concerns the extent of the organizational transformations. They may have been in humans and the more developed mammals limited, partial and non-integrated before they developed into schemata affecting the whole of the cognitive system (or brain, in line with JAYNES' 1976 claim). Another strand concerns control of their evocation. Initially there may have been no possibility for an intentional evocation. One had simply to wait until an organizational transformation of a familiar or nonfamiliar kind would happen. At some point some modicum of control was gained by applying external means, ranging from exposure to specific environmental conditions, hunger, sleeplessness, sensory deprivation, drugs, or particular behavioral or cognitive techniques. Not surprisingly, the special organizational transformations we call SOC remained often under the secret sway of particular people (e.g., Shamans, prophets, priests, soothsayers) who kept the privilege of access to the SOCs because of the power or other advantages it may have given them over others. The next developmental stage has been attained only recently with the possibility of evoking the SOCs by purely cognitive means and at will (e.g., KREITLER/KREITLER/WANOUNOU 1987–88).

Finally, the third evolutionary strand concerns the number and variety of SOCs. Initially there must have been only one or two SOCs, and they were easy to identify and characterize.

In the course of time many more SOC came to be evoked and recognized. The availability of a great number of different SOCs indicates the possibility of adapting to each type of cognitive task the optimal SOC and gaining control over ourselves and over emotions at will by evoking the SOC appropriate for the task at hand. This possibility appears particularly likely in view of the availability of easily attainable cognitive means for controlling the evocation of the SOCs. However these possibilities suggest that the next step in this evolutionary development would be attained when different SOCs would be invented—namely, defined and evoked at will—in view of definite cognitive, emotional and other goals. With the developments taking place in the sphere of virtual reality and the means provided by the system of meaning this stage could not be far ahead.

Author's address

Shulamith Kreitler, Dept. of Psychology, Tel-Aviv University, Tel-Aviv 69978, Israel. Email: krit@netvision.net.il

References

Baars, B. J. (1998) A cognitive theory of consciousness. Cambridge University Press: Cambridge MA.

Baars, B. J. (2001) There are no known differences in brain mechanisms of consciousness between humans and other mammals. Animal Welfare 10:31-40.

Bentall, R. P. (2000) Hallucinatory experiences. In: Cardeña, E./Lynn, S. J./Krippner, S. (eds) Varieties of anomalous experience. APA: Washington DC, pp. 85–120.

de Waal, F. B. M. (1986) Deception in the natural communication of chimpanzees. In: Mitchell, R. W./Thompson, N. S. (eds) Deception: Perspectives on human and non-human deceit. State University of New York Press: Albany NY.

Faber, M. D. (1981) Culture and consciousness: The social meaning of altered awareness. Human Sciences: New York, pp. 221–244.

Fischer, R. (1978) Cartography of conscious states: Integration of east and west. In: Sugerman, A. A./Tarter, R. E. (eds) Expanding dimensions of consciousness. Springer: New York, pp. 24–57.

Jaynes, J. (1976) The origin of consciousness in the breakdown of the bicameral mind. Houghton Mifflin: Boston MA.

Johnston, W. (1971) The still point. Harper & Row: New York.

Kreitler, H./Kreitler, S. (1982) The theory of cognitive orientation: Widening the scope of behavior prediction. In: Maher, B. A./ Maher, W. A. (eds) Progress in experimental personality research, vol. 11. Academic Press: New York, pp. 101–169.

Kreitler, H./Kreitler, S. (1990a) The psychosemantic foundations of creativity. In: Gilhooly, K. J./Keane, M./Logie, R./ Erdos, G. (eds) Lines of thought: Reflections on the psychology of thinking, vol. 2. Wiley: Chichester, pp. 191–201.

Kreitler, S. (1965) Symbolschöpfung und Symbolerfassung: Eine experimentalpsychologische Studie. Reinhardt: Basel/München.

Kreitler, S. (1999) Consciousness and meaning. In: Singer, J. A./Salovey, P. (eds) At play in the fields of consciousness. Erlbaum: Mahwah NJ, pp. 175–206.

Kreitler, S. (2001a) Psychological perspective on virtual reality. In: Riegler, A./Peschl, M. F./Edlinger, K./Fleck, G./Feigl, W. (eds) Virtual reality: Cognitive foundations, technological issues and philosophical implications. Peter Lang: Frankfurt, pp. 33–44.

Kreitler, S. (2001b) An evolutionary perspective on cognitive orientation. Evolution and Cognition 7:81–97.

Kreitler, S. (in press a) Psychosemantics of self and other. Self and Identity.

Kreitler, S. (in press b) Treatment-by-meaning of retarded individuals. Plenum: New York.

Kreitler, S./Kreitler H. (1984) Meaning assignment in perception. In: Fröhlich, W. D./Smith, G. J. W./Draguns, J. G./ Hentschel, U. (eds) Psychological processes in cognition and personality. Hemisphere publishing/McGraw-Hill: Washington DC, pp. 173–191.

Kreitler, S./Kreitler, H. (1985a) The psychosemantic determinants of anxiety: A cognitive approach. In: Van der Ploeg, H./Schwarzer, R./Spielberger, C. D. (eds) Advances in test anxiety research, vol. 4. Swets & Zeitlinger: Lisse, pp. 117–135.

Kreitler, S./Kreitler, H. (1985b) The psychosmenatic foundations of comprehension. Theoretical Linguistics 12:185–195.

Kreitler, S./Kreitler, H. (1986a) Individuality in planning: Meaning patterns of planning styles. International Journal of Psychology 21:565–587.

Kreitler, S./Kreitler, H. (1986b) The psychosemantic structure of narrative. Semiotica 58:217–243.

Kreitler, S./Kreitler H. (1986c) Schizophrenic perception and its psychopathological implications. In: Hentschel, U./ Smith, G./Draguns, I. G. (eds) The roots of perception. North-Holland: Amsterdam, pp. 301–330.

Kreitler, S./Kreitler, H. (1986d) Types of curiosity behaviors and their cognitive determinants. Archives of Psychology 138:233–251.

Kreitler, S./Kreitler, H. (1987a) Modifying anxiety by cognitive means. In: Schwarzer, R./Van der Ploeg, H./Spielberger, C. D. (eds) Advances in test anxiety research, vol. 5. Swets & Zeitlinger: Lisse, pp. 195–211.

Kreitler, S./Kreitler, H. (1987b) The motivational and cognitive determinants of individual planning. Genetic, Social and General Psychology Monographs 113:81–107.

Kreitler, S./Kreitler, H. (1987c) Plans and planning: Their motivational and cognitive antecedents. In: Friedman, S. L./ Scholnick, E. K./Cocking, R. R. (eds) Blueprints for thinking: The role of planning in cognitive development. Cambridge University Press: New York, pp. 110–178.

Kreitler, S./Kreitler, H. (1987d) Psychosemantic aspects of the self. In: Honess, T. M./Yardley, K. M. (eds) Self and identity: Individual change and development. Routledge & Kegan Paul: London, pp. 338–358.

Kreitler, S./Kreitler, H. (1988) Meanings, culture and communication. Journal of Pragmatics 12:135–152.

Kreitler, S./Kreitler, H. (1989) Horizontal decalage: A problem and its resolution. Cognitive Development 4:89–119.

Kreitler, S./Kreitler, H. (1990b) Cognitive foundations of personality traits. Plenum: New York.

Kreitler, S./Kreitler, H. (1991) The circle of meaning dimensions: Applications of Guttman's methodology in the study of cognition. Megamot 33:342–358 (Special issue: 'Facet theory and its applications', devoted to Louis Guttman) (in Hebrew).

Kreitler, S./Kreitler, H. (1993a) The cognitive determinants of defense mechanisms. In: Hentschel, U./Smith, G./Ehlers, W./Draguns, I. G. (eds) The concept of defense mechanisms in contemporary psychology: Theoretical, research and clinical perspectives. Springer: New York, pp. 152–183.

Kreitler, S./Kreitler, H. (1993b) Meaning effects of context. Discourse Processes 16:423–449.

Kreitler, S./Kreitler, H. (1994) Motivational and cognitive determinants of exploration. In: Keller, H./Schneider, H./ Henderson, B. (eds) Curiosity and exploration. Springer: New York, pp. 259–284.

Kreitler, S./Kreitler, H. (1997) The paranoid person: Cognitive motivations and personality traits. European Journal of Personality 11: 101–132.

Kreitler, S./Kreitler, H/Wanounou, V. (1987–1988) Cognitive modification of test performance in schizophrenics and normals. Imagination, Cognition, and Personality 7:227–249.

Lahav, R. (1982) The effects of meaning training on creativity. Unpublished master's thesis. Tel-Aviv University: Tel-Aviv, Israel.

Lorenz, K. (1981) The Foundations of Ethology. Springer: New York.

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

Singer, J. L. (1970) Drives, affect and daydreams: The adap-

tive role of spontaneous imagery or stimulus-independant mentation. In: Antrobus, J. S. (ed) Cognition and affect. Little, Brown: Boston MA, pp. 131–158.

Singer, J. L./Antrobus, J. S. (1972) Daydreaming, imaginal processes, and personality: A normative study. In: Sheehan, P.W. (ed) The function and nature of imagery. Academic Press: New York, pp. 175–202.

von Frisch, K. (1967) The dance language and orientation of bees. Harvard University Press: Cambridge MA.

Wilson, E. O. (1971) The insect societies. Harvard University Press: Cambridge MA.

Wimmer, M. (1995) Evolutionary roots of emotions. Evolution and Cognition 1:38–50.

Wulff, D. M. (2000) Mystical experiences. In: Cardeña, E./Lynn, S. J./Krippner, S. (eds) Varieties of anomalous experience: Examining the scientific evidence. APA: Washington DC, pp. 379–440.

Adolf Heschl

Genes for Learning

Learning Processes as Expression of Preexisting Genetic Information

Abstract

It is commonly assumed that learning in both animals and humans represents a particular, since essentially non-genetic way to assimilate information from the environment. This view is even held by many theoreticians who, otherwise, repeatedly stress the importance of the genetic adaptation process by referring to the general validity of the so-called central dogma of molecular biology which forbids any directed instruction of the genome through phenotypic influences. If, however, one takes a closer look at what really happens in a number of most different learning processes, one quite rapidly discovers that they fully obey the central dogma and, associated with it, the mutationist principle of evolutionary theory which prescribes random variation as the sole source of evolutionary change, i.e., information gain. Consequently, learning taken to be a true gain in information is possible only through concrete genetic changes within the germ line, i.e., through non-somatic mutations, and has nothing to do with our common understanding of "learning" as an ontogenetic phenomenon.

Key words

Central dogma, mutationist principle, information gain, learning, strong genetic principle, behavioral genetics.

Introduction

Usually, evolutionary studies in biology begin with a comparison of morphological structures of different organisms (for the methodology of comparative morphology, see RIEDL 1977), then move on to histological and cellular features and eventually finish with a detailed systematic analysis of molecular components based on sequencing data of both proteins (amino acids) and nucleic acids (nucleotides). This procedure, which to a certain degree reflects the historical development of biology as a scientific discipline, has led to a huge amount of knowledge about the concrete process of evolution on earth, i.e., the specific phylogenetic pathways within and between the different animal classes. What surprises at that procedure, that is the fact that even though the different methodological approaches very often have been opposed to each other as being partially incompatible (cf. PATTERSON 1987; HILLIS 1994), the final result in most cases was a very similar, if not completely identical one:

"When genetic correlations are well estimated, they tend to be not very different in either magnitude or pattern from their phenotypic counterparts. Thus, when reliable genetic estimates are unavailable, phenotypic correlations and scaled variances may be substituted for their genetic counterparts in evolutionary models of phenotypic evolution" (CHEVERUD 1988, p966).

"Hopefully, then, a better understanding of homologous relationships at the molecular level can lead to a better understanding of homology in general" (HILLIS 1994, p360).

This proves that evolutionary biology was always on the search of something very fundamental in living systems and that is the universal factor of *homology* or "true identity", hidden behind the immense degree of phenotypic divergence in the living world. In the words of Richard OWEN, one of the great pioneers in homology research: Identity of descent, that is "the same organ in different animals under every variety of form and function" (OWEN 1843). In principle, already DARWIN's discovery of evolution itself was nothing more than the first and still rough description of what today is much more concretely defined as material, that is genetic relatedness in the double meaning of the word: evolutionary or phylogenetic (cf. "genesis") and genetic in a narrow sense (cf. "genes").

Otherwise, the construction of all those surprisingly stable genealogies would not have been possible, if phenotypic and genotypic characters would

not be correlated with each other in that strong, even though sometimes very indirect causal manner. When we move now from purely morphological characters to what is of particular interest for the chosen subject of the present study, that is behavioral features, we still remain on the phenotypic side which becomes obvious the moment we have a closer look at the subtle transitions between obviously "hard" morphology and seemingly "soft" function, i.e., behavior:

"Broadly speaking, behavior is what animals do. However defined, the distinctions between behavior and other phenotypic attributes are fuzzy. Enzyme-substrate reactions and other fast molecular events are referred to as biochemical or physiological, somewhat slower responses are called behavior, and features that appear stable over long periods are known as morphology. If we repeatedly measure the antlers of deer, the feathers of birds, or the uterine mucosa of mammals with estrous cycles we sample a kind of variability equivalent to repeated expression of a particular behavior… Thus far, behavior does not seem to be especially variable nor especially subject to homoplasy compared to morphology" (GREENE 1994, p378).

Due to the early works of Konrad LORENZ who, together with Karl VON FRISCH and Nikolaas TINBERGEN, was one of the co-founders of classic European ethology, we know today that the above view must be correct. In his famous study from 1941 on the occurrence of certain innate behavior patterns (preening, mating displays, vocalizations) in 18 different species of ducks (*Anatinae*), he succeeded to demonstrate that it is indeed legitimate to transfer the evolutionary idea of common genetic descent, up to then reserved mainly for morphology, to the domain of behavior. At that time, however, only a few researchers followed this promising route (ALEXANDER 1962; ANDREW 1956; CULLEN 1959; MAYR 1958; TINBERGEN 1951, 1959), among others because it was often too laborious to conduct similarly comprehensive studies (LORENZ was, as he described himself, "lazy" enough to hold out).

But later on, as the total sum of data on concrete phylogenetic relationships among different phyla has enormously increased, studies on the evolution of behavior slowly began to spread. And in 1990, a replication of LORENZ' pioneer study was made with the aid of modern computer analyses based on highly developed assumptions about evolutionary rates of change and systematic relationships. The result was an impressive confirmation of the validity of this first rather gestaltist approach (BURGHARDT/GITTLEMAN 1990). Meanwhile, it seems as if a new wave of interest in "good old" comparative ethology has led to a respectable number of phylogenetic investigations with quite remarkable results (GREENE/BURGHARDT 1978; MCLENNAN/BROOKS/MCPHAIL 1988; LOSOS 1990; PRUM 1990; BROOKS/MC LENNAN 1991; EDWARDS/NAEEM 1993; GREENE 1994). And, roughly, all studies come to the same conclusion, a conclusion which was already anticipated some years ago by the evolutionary biologist G. G. SIMPSON (1958, p54): "similarity of behaviours tends, like structural similarity, to be proportional to phylogenetic affinity". Hence, for the case of proved instincts, evolutionary theory has revealed to be a very useful scientific concept, that means a concept which is clearly testable in an empirical way.

If, however, one is seriously interested in dealing with the question of the evolution of learning, one finds oneself confronted with a completely different and rather difficult situation: All phylogenetic studies so far done in the behavioral sciences were limited to so-called innate or, in more modern terms, "genetically programmed" behavior patterns since the proof of genetic inheritance has always remained the necessary precondition for any phylogenetic comparison. What then about the relationship between learning and evolution, if the first one is defined by the exact opposite to instinctive behavior, i.e., the complete lack of a provable genetic causation? Even though most of the modern theorists of behavior vigorously contest any usefulness of the traditional innate/acquired-dichotomy (compare for example PLOTKIN 1994, p165: "The doctrine of separate determination is completely wrong" with MARLER 1991, p43: "Such dichotomising is not merely unfruitful but logically incorrect"), it nevertheless remains obvious that the underlying distinction is far from having disappeared. This may even be the case for purely pragmatic reasons, as SHETTLEWORTH (1998, p16) recently has brought it to the point: "we do sometimes need a term for the many behaviors that appear in development ready to serve their apparent function before they have done so".

Basically, there are three important possibilities to solve this dilemma, depending on how close one positions the phenomenon of learning with regard to biology. The first one is exemplified by a recent review on the field published by the behaviorist M. E. BITTERMAN (2000), in which he presents to us, as he specifies, "a psychological perspective" on "cognitive evolution". In this approach, he defends general-process theory from human psychology to be still powerful enough to encompass all known learn-

ing processes in most animals, from comparatively primitive insects like honeybees up to higher vertebrates. This is indeed possible if one proceeds from the assumption that something like *general laws of learning* are active in all species with only minor quantitative deviations from one and the same abstract relationship as for example assumed for the case of PAVLOVIAN conditioning processes, where "the constants may vary widely in value… from species to species, but the learning equation may be the same" (BITTERMAN 2000, p64; RESCORLA/WAGNER 1972):

$$\Delta V = \alpha \cdot U\beta \cdot (\lambda - V)$$

with V = strength of association between unconditioned (US) and conditioned stimulus (CS) at the beginning of each trial, α = salience of CS, $U\beta$ = learning rate, λ = maximal strength of association.

However, in the sequel the connection with real biological evolution must rapidly go down, because, as MACPHAIL (1982) has successfully demonstrated already some years ago, no relevant qualitative differences are to be detected if one makes similar comparisons of species on the basis of such a purely functionalist approach. And, in some way, this outcome is even to be expected. If, as a behavioral biologist, one would choose for example a universal ability to associate as a valid taxonomic character (which however it isn't), then of course the resulting phylogenetic tree must necessarily shrink to a quite undifferentiated stump which, in the meantime, has become famous as MACPHAIL's "null hypothesis" (1985).

In contrast to BITTERMAN´s review, which does not really go beyond the traditional behaviorist paradigm, the new textbook written by Sarah SHETTLEWORTH on the subject of "Cognition, Evolution, and Behavior" (1998) represents already a much more straightforward approach to a newly emerging biology of learning, in the sense that she tries therein to keep some sort of balance between the narrower psychological and the proper evolutionary approach. Hence, it is fully accepted that genes, which are not a really interesting subject of investigation for the majority of psychologists, are at least very important prerequisites for every category of behaviour. But in the end, it is also retained that genes can never completely determine the concrete structure of a cognitive apparatus, with the exception of a few provedly innate reflexes or fixed action patterns (e.g., display behavior). Hence, SHETTLEWORTH basically agrees with the idea of a phylogeny of learning mechanisms by clearly voting for a system of innately pre-given cognitive modules as is currently much advocated in so-called evolutionary psychology (COSMIDES/TOOBY 1994). Interestingly enough, the concept of modules quite narrowly resembles the meanwhile outdated instinct concept from classical ethology (LORENZ 1937; TINBERGEN 1942), but has comparatively acquired a much greater acceptance because of its apparently much more technical background.

A third approach, which, on the one hand, adopts the valuable insights of the second, compromise-type one, but at the same time is placed even stronger on the biological side, will be elaborated in what follows. Then, a discussion of the empirical situation in behavioral genetics is added to better evaluate the new perspectives of a primarily genetic approach. Finally, some potentially fruitful possibilities of a closer future cooperation between learning research and genetics are outlined.

Evolutionary Theory and Learning

Basically, it are two important axioms from evolutionary theory which suggest a much stronger dependence of learning behavior from pre-existing genetic information than assumed till now. The first one is the so-called central dogma of molecular biology which, as such, strictly forbids any informational influences from the phenotype (proteins) back to the genotype (nucleic acids):

$$\text{DNA} \Leftrightarrow \text{RNA} \Rightarrow \text{protein}$$

Historically, this axiom goes back to the founder of modern molecular genetics, August WEISMANN who was the first to show that this principle is also valid for the relationship between ontogeny and phylogeny (WEISMANN 1892). In his view, no influence from the developing soma is able to change in any directed way the information stored in the cells of the germ line (WEISMANN's doctrine). The empirical refutation of Lamarckism in the course of the 19th and 20th century showed that he was basically right with that postulate: Neither cutting off thousands of mouse tails nor mutilating an equally large number of amphibians ever led to a significant change in the animals' morphology.

At first sight, the central dogma seems to have nothing to do with the question of learning. However, learning is just defined by the assumption that the organism, by way of its realization, acquires new knowledge about the surrounding world. Now, if we agree with this view—and the

current status quo in learning research undoubtedly does so—, which is the material medium for the storage of this newly acquired knowledge? Of course, if by definition the medium is not the genotype, only a phenotypic substrate can come into question for fulfilling this purpose. Anyhow, following the central dogma, all phenotypic characters must be assumed to be well-controlled end products of the genes, which, in a first step, code only for proteins, but indirectly, i.e., via a variety of interacting molecular networks also determine the structure of all other physiologically important structures. Hence learning should be able to induce the occurrence of basically new, i.e., *novel* forms of phenotypic structures and this in a way completely independent of the underlying genetic system, something, which has never been observed in any living system. The conclusion we have to draw from this basic asymmetric relationship says that learning must be defined otherwise than by the idea that it is a different, since non-genetic way to assimilate information from the environment. Consequently, the traditional view could be called the LAMARCKIAN interpretation of learning.

The second axiom is closely interconnected with the first one in a sense because, in principle, it merely postulates theoretically what has been already established by the central dogma in a purely experimental way. It concerns the mutationist principle of biological evolution which, in a nut-shell, states that only true random changes can lead to new acquisitions, depending on their success in DARWINIAN struggle for life, i.e., natural selection. Random genetic changes, i.e., mutations do fulfil this precondition, but so do not any phenotypic structures which always remain causally dependent from pre-existing genetic information. At least, as far as we know, no phenotypic character is able to spontaneously mutate in a truly random manner. Neither proteins nor other cellular substances are known to change their composition or structure in this way to perhaps form the basis for the acquisition of new elements of information (e.g., new amino-acid sequences). At the same time, learning is often interpreted as representing a goal-directed and hence regular form of acquisition of new information (cf. "laws" of association, learning "mechanism" etc.), but this idea too, if taken to be true, would violate the basic principle of random variation. In this way, the mutationist principle of evolutionary change becomes one of the strongest arguments against the very common identification of learning with an increase in the informational status of the organism (cf. HESCHL 1990).

Some more technical arguments against the conventional learning paradigm concern the concrete modes of transition from purely genetic to the first really non-genetic types of behavior, a problem which imposes itself both with respect to phylogeny and ontogeny. The respective questions in short: 1. If evolution began with genetically determined behavior patterns—and we must necessarily assume that this was indeed the case after the origin of multicellular organisms—, then how was it possible to switch in a purely *genetic* way from still plain genetic determination over to a first non-genetic indeterminism? 2. If the very first behavior patterns shown by young animals (man included) are innate (e.g., reflexes, fixed action patterns)—and it is commonly assumed that this is in fact the case—, then again how can it be possible to realize the same mysterious transition during ontogeny, namely from full determination by genes to a sudden non-genetic liberty of the individual? These questions, as inextricable as they may appear at first sight, are comparably easy to answer if one dispenses in advance with the postulate of the real existence of such a transition. Empirical research, at least as far it concerns the concrete details of much more accessible ontogeny (embryogenesis, morphological and behavioral development), confirms this idea: There exists not a single proof of such a categorical transition (BROWDER 1984).

Last but not least, we have to briefly discuss the meanwhile quite common picture of the learning process as being represented by some sort of a frame (2D metaphor) or container (3D metaphor), which can be filled up with a certain amount of information stemming from the environment. Thereby, the concrete dimensions of such a container are defined by the limits or "biological constraints" of the learning mechanism, whereas the content itself basically is thought to remain independent of these limitations. Learning in that perspective equals a refinement or, better: fine-tuning of some pre-existing, but yet too coarse cognitive structures to deal successfully with the environment. Irrespective of the above-mentioned, more general objections to the validity of this metaphor, if we accept this view we are automatically confronted with the severe problem of interpreting the seemingly structure-less content of such a container as a representation of the assumed result of a learning process and, that is always *well-structured* information about the environment.

By summarizing this short excursion into evolutionary theory, we are now able to postulate the

validity of a so-called *Strong Genetic Principle* (cf. MADDOX 1993) also for the disputed case of learning behavior (cf. HESCHL 1994, 1996). This implies that, in principle, behavior patterns which are described in learning research can be treated like innate instincts, even though they are equipped with both higher degrees of complexity and flexibility. Learning, in such a view, is identical with the ontogenetic expression of pre-existing genetic programs, which determine also the very details of this kind of behavior. In the end, learning will even completely lose its traditional status, namely to have anything to do with the acquisition of new information.

Empirical Evidence

It is one thing to deduce some logical conclusions from a general theory such as the theory of evolution, but it is a different thing to demonstrate that these same conclusions can also withstand severe empirical testing. Now, it is undoubtedly by far to early to pretend that the current situation in biological research about the detailed molecular causation of behavior and especially learning is already advanced enough to be able to deliver a comprehensive proof of what has been asserted in the introductory sections before. But nevertheless, there exist already some very interesting empirical hints, which at least roughly indicate the approximate direction, in which future studies in the field probably may develop. To better understand what is meant by that, let us consider a series of quite recent research results which, slowly and step by step, come closer to our actual subject of interest.

Genes coding for the formation and functioning of a complete nervous system

In a review article from 1994, James THOMAS, a geneticist from the University of Washington, gave the first molecular description of what he calls the "Mind of a Worm" (THOMAS 1994). In this report, he nicely demonstrates the enormous amount of currently existing knowledge about both the organization and functioning of the CNS of a primitive multicellular organism. *Caenorhabditis elegans*, a small soil nematode, is characterized by its miniature form of a nervous system which consists of about 300 neurons, 5000 chemical synapses, 600 gap junctions and 2000 neuromuscular junctions.

With the now very advanced technique of killing only a small set of neurons by using a laser microbeam, no less than 40 different classes of neurons could be discriminated which are assumed to play a specific behavioral role in the animal. In addition, one meanwhile knows the exact cell lineages that are associated with the developmental origin of each neuron (SULSTON et al. 1983), all the main synapses between these neurons (from serial section electron micrographs) and, since only recently, the complete genome sequence (HODGKIN et al. 1998), which is required to build not only the CNS, containing about one-third of all extant somatic cells (BARGMANN 1998), but also the animal as a whole (RUVKUN/HOBERT 1998).

The analysis, however, has not been stopped at the pure description of the relationship between genes and the structure of this mini-brain. Meanwhile, more than 250 genes have already been identified to regulate in a specific way the worm's behavior. To give only one example of a seemingly rather simple behavior: defecation, which after all consists of a well-tuned spatiotemporal coordination of three separate motor patterns, namely posterior body muscle contraction, anterior body muscle contraction, and expulsion or enteric muscle contraction, is controlled by a series of different genes. This could be convincingly demonstrated by the proof of 24 concrete behavioral mutations which either changed only single steps within the whole process without perturbing the remaining ones or influenced the coordination between different phases of the behavior (THOMAS 1990).

Genes coding for the sensitivity of sensory organs

With regard to sensory capabilities, the performance of the nematode's brain, a simple nerve ring around the pharynx, is even more astounding. *C. elegans* is able to distinguish between seven classes of different odorants with the help of 15 different classes of chemosensory neurons. This explains how it can be possible that the worm perceives various gradients of at least 120 different odorants. More than 20 genes exist which have been revealed to influence these quite specific sensory reactions, but at the same time mutations with often very complex patterns of odorant response defects have been found. Meanwhile, a new analysis of the chemosensory system has documented the existence of over 600 genes, which encode only chemoreceptors (TROEMEL et al. 1995) and it is still widely unknown, how this complex genetic information is transferred into the functioning of the respective neurons.

Genes coding for the establishment of simple behavioral associations

Even though *C. elegans* is held to be a still rather primitive creature which of course lack the behavioral complexity and plasticity that is characteristic for higher animals, this tiny worm nevertheless shows several forms of nonassociative learning the result of which can persist up to a day. Even the possibility of true associative learning has been proved by a quite recent study conducted by WEN et al. (1997) who also succeeded in detecting two genetic mutants which both are impaired in associative learning. At the same time, these mutations didn't alter normal nonassociative learning and sensorimotor function, which suggests that there exists a separate path of determination. Meanwhile, however, already over 300 genes are known that code for a variety of different protein kinases, which are suspected to play an important role in the regulation of the worm's more flexible behavior patterns (BARGMANN 1998). In fact, behavioral flexibility in *C. elegans* has been stated in the area of thermotactic, chemotactic, and mechanosensory behavior (COLBERT/BARGMANN 1997).

The so far best known case, where genes affect quite directly associative learning processes are diverse mutants in the fruitfly *Drosophila melanogaster*. Flies of that species have been exposed to an odour cue and simultaneously given an electric shock. The animals soon learned to avoid the conditioned cue in the absence of a shock. Usually, this avoidance response was kept for several hours afterwards, but learning mutants lost this ability quite rapidly, which manifested itself in a completely different retention curve. This was particularly obvious in *dunce* mutations of *D. melanogaster*, which, in addition, never did reach the normal learning level of unchanged individuals (TULLY/QUINN 1985).

The discovery of *dunce* and some other behavioral mutations allowed a first investigation of the concrete biochemical processes in learning retention in *Drosophila*, which seems to be controlled by the influence of a phosphodiesterase enzyme responsible for breaking down compound cyclic adenosinemonophosphate (cAMP). In the mutant *D. melanogaster dunce*, this enzyme is not active, which results in elevated levels of cAMP (DAVIS/DAUWALDER 1991). Meanwhile, high activity of phosphodiesterase has been proven to be important for neurotransmission, particularly in structures known as mushroom bodies, which, in insects (e.g., bees), are known to play a key role in the learning process.

One could now argue that the relationship between the *dunce* locus and learning is rather a very unspecific one, determining only in a very general manner the biochemical basis for the actual so-called "information storing processes". That this cannot be the case becomes clear in a more recent study, in which another mutation, *Volado*, has been shown to be responsible for the formation of short-term memory in certain mushroom body (neuropil) cells (GROTEWIEL et al. 1998). *Volado* is acting via the encoding of integrin, a molecule, which is required for the maintenance of olfactory memory during the first 3 minutes of the training. Many more additional loci are expected to influence also other details of the neuronal association procedure (DUDAI 1988). This is not surprising, if again we just bear in mind the already considerable number of genes (approx. 1220; BARGMANN 1998), which alone were found to code in the comparatively primitive nematode *C. elegans* for neuronal functions.

Genes coding for stimulus-specific habituation and "forgetting" processes

Zebra finches (*Taeniopygia guttata*) were shown to habituate to repeated, unreinforced presentations of complex sounds by a decrease of neuronal responses in the caudomedial neostriatum (NCM) (CHEW/VICARIO/NOTTEBOHM 1996). After successful habituation, the animals slowly "forgot" about the learning in dependence of the time elapsed since the training. So far, this is still commonplace in conventional learning research. Then, however, things began to become more interesting. First, Chew and colleagues detected specific time periods, namely six in number, during which some kind of a spontaneous forgetting process occured, determined alone by stimulus class, number of presentations, and interval between presentations (2–3, 6–7, 14–15, 17–18.5, 46–48, 85–89 hours after first exposure to stimulus). Accordingly, the researchers termed their discovery "quantal duration of auditory memories" and they observed that, as was to be expected, the longest memory durations are shown for songs of conspecifics and, among them, female calls.

For the present subject matter, however, the most interesting result in this study comes from an analysis of gene expression and protein synthesis patterns in the corresponding brain area, i.e., the NCM. It turned out that at least the first five periods of forgetting are correlated with periods of gene expression and subsequent protein synthesis involved in the maintenance of the longer lasting (85–89 hrs.)

habituation. Hence, it seems as if genetic influence is both permanently accompanying as well as guiding these kinds of learning processes and the study conducted by CHEW, VICARIO and NOTTEBOHM is undoubtedly one of the most convincing proofs currently available of that relationship.

Genes coding for complex learning and memory

Behavioral genetics in mammals is currently found in extremely fast development, although, compared with the situation in nematodes and insects, it is by far not as easy to isolate appropriate mutants. But, in some sense, this is to be expected. The more complex the morphology and physiology of an organism is, the more difficult it necessarily must be to decipher the intricate network of genetic instructions, which make that organism working. The most promising methods so far applied in single-gene studies contain both "forward" and "reverse" genetics approaches. Whereas the first, more conventional approach goes from rather rough methods of genetic manipulation to the phenotype by applying classical mutagenesis (agents: X-rays, ENU, Chlorambucil, transgenes, gene traps) and screening procedures (behavioral tests), the latter takes exactly the opposite way: "back" from a new and yet unknown phenotype to the supposed genes responsible for it. Today, reverse genetics is very successfully done by replacing well-defined genes of choice by a range of mutated variants (base substitutions or deletions: null mutations, knockouts etc.) which are then introduced into the respective organism´s germ line through a new method called gene targeting. Since only a few years, both techniques, which initially have been developed mainly in *Drosophila*, can also be applied to the analysis of the molecular mechanisms underlying mouse behavior (TAKAHASHI/PINTO/VITATERNA 1994).

This impressive progress in methodology has made it possible that, meanwhile, the first empirical hints towards a genetic influence on explicit learning come from the mouse. Until now, we discussed only implicit forms of learning, i.e., those kinds of conditioning, which are characterized by the establishment of associations that do not necessarily include retrievable sensory (and hence potentially conscious) representations. In the case of simple spatial or contextual learning, the situation is somewhat different in the sense that we can assume the formation of certain limited learning contents in the animal, which are associated with the conditions of acquisition. This new level of behavior corresponds to what is quite roughly called "memory" or, in neurobiological terms, synaptic long-term potentiation (LTP), the latter being able to permanently change existing excitabilities at the junctions between neurons.

Now, even this kind of already quite sophisticated learning behavior has revealed to be not independent of strong genetic influences: in mice, at least 8 mutants (*Prn-p, Camk2a, fyn, src, Pkcc, Syn-1, Ncam, Nos-2*) are already known, which exhibit marked differences in their capacity for spatial learning (learning task: Morris water maze; review in TAKAHASHI/PINTO/VITATERNA 1994). This result, however, must still be taken as a very preliminary one, since most complex behaviors, especially among vertebrates, are generally assumed to be under—equally complex—polygenic control. Moving then to our own species, *Homo sapiens*, we will see that things become even more complicated.

Genes coding for symbolic communication

In general, language is held to be that medium, which can be interpreted as being much more independent of the influence of specific genetic instructions than any other type of human behavior. Indeed, the case is rather clear. Language in the young child is learned by imitation of other people, usually adult family members who, by speaking to their after-growth, make sure that the so-called "mother tongue" of the population, they are living in, is transferred from one generation to the next one. This seemingly simple procedure is thought to form the basis of what we call human culture and tradition, a phenomenon which even defines itself by the strict exclusion of biology.

However, the meanwhile famous case of the family F from Great Britain, which has been thoroughly investigated by the linguist Myrna GOPNIK, raises serious doubts about this seemingly coherent picture. To be brief, let us take a short look at probably the first true genealogical and at the same time also phylogenetic tree (1st and 2nd order) on a purely linguistic basis. If we define dysphasia d (deficiency in applying general grammatical rules; e.g. tense: "Every day, he kisses his nanny. Yesterday, he ???", plural: "book—books. wug—???") as a relevant taxonomic trait, we arrive at the following situation depicted in Table 1.

The conclusion to be drawn from that unique case study, is a much straightforward one. The fact that the grandchildren showing dysphasia are distributed in a completely random manner among families, sexes and sibling constellations makes it highly

Grandmother & Grandfather ↓				
Daughter 1 ↓	*Daughter 2* ↓	*Daughter 3* ↓	*Son 1* ↓	Son 2 ↓
m (17)	f (21)	f (17)	m (10)	m (20)
m (15)	f (17)	*m (16)*	f (07)	f (18)
m (14)	*f (12)*	f (12)	*f (05)*	
f (12)	*f (11)*	*m (08)*	*f (02)*	
m (10)	*m (07)*			
m (08)				
f (07)				
m (06)	fraternal			
f (06)	twins			

Table 1: Subjects in *italic* were those diagnosed as dysphasic, m = grandson, f = granddaughter; age in parentheses. From GOPNIK 1990.

probable that there exists a single dominant gene, which seems to be responsible for the observed pattern of inheritance. This does not necessarily mean, however, that we have detected the "gene for grammar", but it says at least that there is in fact a gene, which plays a very important role for the development of grammatically normal speech in humans.

At the same time, the effect has revealed to be a quite specific one, since serious dysphasia like that observed in the mentioned family did not influence other, more general cognitive capacities. In fact, all family members were found to lay within the normal range of variation in intelligence, i.e., more or less closely around the statistical mean of the whole population. In addition, only one particular grammatic faculty was impaired and that was the accurate usage of certain syntactical–semantic rules in language (number, gender, animacy, proper names, tense, aspect), whereas other linguistic skills with comparable complexity remained unchanged.

In other words, already the early PIAGET (1945) was right when he postulated a basic independence of intelligence from speech (but a strong causal dependence the other way around) even though he would certainly not have much appreciated the idea of genes for language. However, many more additional genes must be assumed to be likely involved in the genetic encoding of language, but this insight is already quite trivial if we remember again the number of 600 genes, which alone code for the diverse chemoreceptors of *Caenorhabditis* (see above).

Genes coding for social intelligence

Social intelligence and normal, i.e., physical or technical intelligence, as measured for example by IQ, are to a certain degree separate aspects of our inherited character. At least, this is assumed in so-called social theory of intellect, which, among others, is grounded on the observation that the evolution of sociality in primates seems to have preceded their particular cognitive capabilities (JOLLY 1966).

Empirical support for this idea comes now from genetic research on Turner's syndrome, a sporadic psychic disorder restricted to human females in which one X chromosome is partly or totally deleted. Interestingly, intelligence of those females concerned usually remains normal, whereas diverse social adjustment problems are quite common. A few years ago, a group of behavioral geneticists could prove that there exists indeed a specific genetic locus for social cognition on the X chromosome (SKUSE et al. 1997). This has been demonstrated by comparing patients with a maternally derived X chromosome (X_m) with patients with a paternal origin of the X chromosome (X_p). The comparison showed that the first group was significantly more severely affected by the loss than the second group, because in the case of maternal X_m the relevant locus is known to be inactivated by genetic imprinting, an intricate mechanism of genetic regulation between generations found only in mammals (REIK/SURANI 1997). Once again, this is just the beginning of more refined molecular genetic research in this area, but it shows already how far it can climb the ladder of complexity.

Genes coding for differences in general cognitive performance (IQ)

Last but not least, we have to address the ever controversial debate about the genetic determination of human intelligence in general. Up to this day, purely phenotypic studies have always suffered from the problem of only theoretically deducing a supposed genetic influence without at the same time having been able to really prove it. The famous "Bell Curve" (HERRNSTEIN/MURRAY 1994) and similar studies (e.g., BOUCHARD 1990) are no exception to that basic methodological restriction. In fact, since the very first vague attempts made by Francis GALTON (1889) most research in so-called "behavioral genetics" as a special subdiscipline of experimental psychology was not true genetic or even molecular research. Instead, it was a rather quite circumstantial procedure of indirectly computing the exact

amount of a so-called "heritability" of behaviors, without however any precise ideas about its concrete relationships to real biological inheritance where, usually, genes are transmitted materially, that means as a whole or not at all, but never to a degree of 50% or so.

But since a few years the field seems to change its ideological orientation. Richard PLOMIN, one of its representatives, has just published a study, in which he begins to call things things (CHORNEY et al. 1998). Starting point for doing so was the insight that, although for years intensive twin and adoption research has repeatedly demonstrated relatively high heritabilities for IQ (more than 50%), "only finding specific genes will convince" (Science 1998) the remaining host of critics. PLOMIN compared the DNA from two groups of children, one with a mean IQ of 103 and another with an averaged IQ of 136. He then applied quantitative trait loci (QTL) analysis to his data, a method which is characterised by the search for a whole set of genes that together govern the quantitative expression of a phenotypic characteristic (details in EDWARDS/STUBER/WENDEL 1987). After having tested for 37 different genetic markers, PLOMIN succeeded to identify a locus in *IGF2R*, an insulin-like growth factor receptor gene on human chromosome No. 6, which exists in two versions, once as allele 4 and once as allele 5.

Then, the phenotypes of the groups were added to the analysis. The group of children with higher IQ turned out to possess at least one copy of allele 5, a rate which was significantly, namely twice as high as in the normal comparison group. Converted into IQ points, this difference amounts to about 2% of the total IQ variance and hence corresponds to about 4 IQ points. Thereby, *IGF2R* itself is not thought to be responsible for the effect, but a functional gene very close to this marker. It is also clear that this gene can only account for a tiny fraction of the entire cognitive endowment, given the complex structure and functioning of the human brain. But nevertheless, PLOMIN's study has nicely documented that the manifold biochemical pathways between genes and the learning brain are basically accessible to scientific investigation, and that, in the near future, we have to expect many more similar results. Finally, the report published in *Science* also cites Nathan BRODY, an American psychologist who freely reflects upon the conceptual background, which ultimately could explain PLOMIN's unexpected results: "There are not even any real theories about what are the biological influences on intelligence." Such a profound scepticism may sound interesting, but here BRODY is definitely wrong, because as we have seen evolutionary theory, if applied in a consistent manner to behavior, just predicts (cf. HESCHL 1990) what only now is slowly becoming to be proved.

"Blind Spots" in Learning Research

The main result of the previous section is not a trivial one. We observed concrete genes, which influence learning behavior in an often very specific way. This is not trivial for the reason that, up to this day, the relationship between learning and biology has always been negatively defined by a complete lack of any specific genetic determination. Likewise, however, this makes it now much more difficult to argue, as has been commonly done, that the stated genetic influence is only a very general biological one and does not touch the essence of learning. The point is rather to understand that even when we go into the most elaborated details of any learning process, we need to explain *why* an animal behaves the way it does and not differently, and, in particular, from where the *information* may stem which is used by the respective behavior. In principle, the situation of the learning animal can be compared with some sort of a permanent decision procedure where the animal is continuously forced to decide upon which direction to choose. At all these decision points, an omnipresent genetic influence is active in every single cell of the body and shows that it is only the animal itself and not the physical influence of the environment which makes the choice. Since, if it would be the latter—as is implicitly postulated by the current paradigm in learning research—no true failure should ever occur. To the contrary, if this would be possible it would always be much easier for the organism to let himself be instructed by the environment about how to solve any vital problem.

Once again, we find ourselves confronted with the still unresolved controversy between the neo-DARWINIAN picture of evolution through random mutation and natural selection and the more popular LAMARCKIAN version of evolution through directed acquisition (and possibly subsequent inheritance) of new characters. It is important for the present discussion to understand that it was exactly this misleading interpretation of learning as being the sole acceptable LAMARCKIAN mechanism (cf. MAYNARD-SMITH 1989, p12: "Given sufficient capacity for learning…, a population can adapt to its environment by non-genetic means"), which was the main reason why its integration into evolutionary theory has been blocked for such a long time. Only today,

the continuously increasing amount of empirical results from behavioral genetics slowly begins to point toward a completely different picture of learning.

But what about the numerous interesting results from learning research itself, shouldn't they be able by themselves to support a stronger phylogenetic approach to the subject matter? I think they could perfectly well, if more researchers in the area would take the decision of concentrating their work much more than hitherto on the investigation of the many remaining true blind spots in learning research, as for example the many cases where an animal, be it a single species or a concrete individual, has been proved to be in fact *unable* to solve a particular cognitive problem. At the same time, learning research could focus in a much more concentrated manner on the properly interesting, that are the real homological and hence genetically related structures in learning behavior. This will be a vast project for the future because the majority of the current experimental paradigms still tacitly favor the behaviorist approach of general-process theory (see introduction), in which a small set of abstract rules of association reduces the enormous amount of diversity of actually existing species specific behavior patterns. Exactly the same would happen (and in earlier times very often happened) in systematic biology if a researcher would hit on the idea to construct a phylogenetic tree by applying purely functional criteria. To give a simple example: Flying is a quite common behavioral feature among animals and we observe this capacity in such different organisms like bees, dinosaurs (e.g., *Rhamphorhynchus*, HOLST 1957), birds, bats and even humans (HOLST 1948). Now, as it is custom in general-process theory, it would be comparably easy to formulate something equally abstract like a universally valid "law of flying", which would allow a perfect quantitative measurement of the phenomena to be investigated (equation of KUTTA-SHUKOWSKI):

$$F \approx p v_0 l Z$$

with F = aerodynamic lift, p = density of medium, v_0 = flying speed (flow), l = wing length, Z = air circulation around wing.

With the results so obtained, we could even construct a perfect empirical confirmation of another null hypothesis (see above) such like: "bees, dinosaurs, birds, bats and humans fly all in a comparable, i.e., qualitatively not discernible manner". Nevertheless, I don't think that we would be well advised to expect that many other evolutionary biologists would agree with us to have conducted an important phylogenetic investigation. To the contrary, for a serious evolutionary taxonomy bees and birds will still remain quite different organisms because—and that is a crucial point—it is always the *totality* of phenotypic characters which must be taken into account before one can try to draft a reliable phylogenetic tree. To come back to flying: It is obvious that the mistake we would have made simply consisted of having confused a merely superficial analogy produced by ecological convergence (convergent selection pressure caused by the medium air) with the reconstruction of the evolution of an intricate network of interrelated homologous behavior patterns.

The best or, at least, most promising way one can see so far to attack this difficult task will be to make a much stronger connection between empirical learning research and already existing taxonomic investigations in comparative ethology and morphology, in particular if the latter are complemented by the newest genetic techniques of calculating phylogenetic trees. Such a procedure is to recommend just because learning of any kind must *always* be linked to some more primitive, that means phylogenetically preceding behavior patterns (e.g., unconditioned stimuli).

Promising Perspectives

The purpose of this article was to demonstrate that molecular behavioral genetics is on the best way to prove that there are no remaining causal "gaps" in the biological study of learning. This explains why the sometimes very long way from genes to behavior must be a clearly deterministic one, a finding, however, which does not exclude complex temporal dynamisms like those found for example in many neuronal calibration processes (cf. KATZ/SHATZ 1998). In such a somewhat altered perspective, the role of the environment is still important, but not as an instructive one as still erroneously assumed in the commonly accepted LAMARCKIAN interpretation of learning. Consequently, we are now approaching a position which, in the end, should confirm the general validity of the central dogma of molecular biology for all kinds of behavior patterns, from very simple conditioning phenomena in invertebrates up to the most sophisticated cognitive abilities in man, including both mental representation (thinking) and symbolic communication (language). In conclusion, all these new results will contribute to a first really unified evolutionary approach to behavior.

Up to that time, however, we have to freely admit that there remain of course many still unresolved questions inherent to the genetic approach, but at the same time it has become clear that it is important to understand what an enormous potential for further research is awaiting the field. Be that as it may, many behavioral geneticists already know which way to choose (note: in what follows, *C. elegans* has been replaced by *Any species*):

"In the long term, the genome sequence points to the need for better behavioral and electrophysiological assays in *Any species* neurobiology. Most mutant screens conducted in the past have required a substantial defect in neuromuscular function such as uncoordinated movement. These screens revealed genes with widespread function and even weak mutations in lethal genes, but they overlooked most genes with subtle, modulatory, or cell-specific functions (note: e.g., involved in learning). The limiting steps for understanding gene function are defining the function of each neuron (still unknown for many *Any species* neurons) and devising better assays for neuronal (note: and behavioral) function in vivo. In the end, instead of transcending neurobiology (note: and psychology), the genome leads back to it" (BARGMANN 1998, p2033).

"Thus, behavioral genetics has entered a new era: it is now possible to study the 'genomics of behavior'" (TAKAHASHI/PINTO/VITATERNA 1994, p1732).

Author's address

Konrad Lorenz Institute for Evolution and Cognition Research, Adolf Lorenz Gasse 2, A–3422 Altenberg, Austria
Email: adolf.heschl@kla.univie.ac.at

References

Alexander, R. D. (1962) The role of behavioral study in cricket classification. Systematic Zoology 11:53–72.

Andrew, R. J. (1956) Intention movements of flight in certain passerines, and their use in systematics. British Journal of Animal Behaviour 4:85–91.

Bargmann, C. I. (1998) Neurobiology of the Caenorhabditis elegans genome. Science 282:2028–2033.

Bitterman, M. E. (2000) Cognitive evolution: A psychological perspective. In: Heyes, C./Huber, L. (eds) The Evolution of Cognition. MIT Press: Cambridge MA, pp. 61–79.

Bouchard, T. J. et al. (1990) Sources of human psychological differences: The minnesota study of twins reared apart. Science 250:223–228.

Brooks, D. R./Mc Lennan, D. A. (1991) Phylogeny, ecology, and behavior. University of Chicago Press: Chicago.

Browder, L. W. (1984) Developmental biology. Saunders: Philadelphia.

Burghardt, G. M./Gittleman, J. L. (1990) Comparative behavior and phylogenetic analysis: New wine, old bottles. In: Bekoff, M./Jamieson, D. (eds) Interpretation and Explanation in the Study of Animal Behavior. Westview Press: Boulder, pp. 192–225.

Cheverud, J. M. (1988) A comparison of genetic and phenotypic correlations. Evolution 42:958–968.

Chew, S. J./Vicario, D. S./Nottebohm, F. (1996) Quantal duration of auditory memories. Science 274:1909–1914.

Chorney, M.J./Chorney, K./Seese, N./Owen, M.J./Daniels, J./McGuffin, P./Thompson, L.A./Detterman, D.K./Benbow, C./Lubinski, D./Eley, T./Plomin, R. (1998) A quantitative trait locus associated with cognitive ability in children. Psychological Science 9:159–166.

Colbert, H. A./Bargmann, C. I. (1997) Environmental signals modulate olfactory acuity, discrimination, and memory in Caenorhabditis elegans. Learning and Memory 4(2):179–191.

Cosmides, L./Tooby, J. (1994) Origins of domain specificity: The evolution of functional organization. In: Hirschfeld, L. A./Gelman, S. A. (eds) Mapping the mind. Cambridge University Press: Cambridge, pp. 85–116.

Cullen, J. M. (1959) Behaviour as a help in taxonomy. In: Cain, A. J. (ed) Function and taxonomic importance. The Systematics Association: London, pp. 131–140.

Davis, R. L./Dauwalder, B. (1991) The Drosophila dunce locus. Trends in Genetics 7:224–229.

Dudai, Y. (1988) Neurogenetic dissection of learning and short-term memory in Drosophila. Annual Review of Neurosciences 11:537–563.

Edwards, M. D./Stuber, C. W./Wendel, J. F. (1987) Molecular-marker facilitated investigations of quantitative-trait loci in maize. I. Numbers, genomic distribution and types of gene action. Genetics 116:113–125.

Edwards, S. V./Naeem, S. (1993) The phylogenetic component of cooperative breeding in perching birds. American Naturalist 141:754–789.

Galton, F. (1889) Natural inheritance. Macmillan: London.

Gopnik, M. (1990) Dysphasia in an extended family. Nature 344:715.

Greene, H. W./Burghardt, G. M. (1978) Behavior and phylogeny: Constriction in ancient and modern snakes. Science 200:74–77.

Greene, H. W. (1994) Homology and behavioral repertoires. In: Hall, B. K. (ed) Homology: The hierarchical basis of comparative biology. Academic Press: San Diego, pp. 369–391.

Grotewiel, M. S./Beck, C. D. O./Wu, K./Zhu, X./Davis, R. L. (1998) Integrin-mediated short-term memory in Drosophila. Nature 391:455–460.

Herrnstein, R. J./Murray, C. (1994) The Bell curve: Intelligence and class structure in american life. Free Press: New York.

Heschl, A. (1990) L = C: A simple equation with astonishing consequences. Journal of Theoretical Biology 145:13–40.

Heschl, A. (1994) Nature/nurture. Nature 369, 185.

Heschl, A. (1996) Biological determinism. Science 271:743.

Hillis, D. M. (1994) Homology in molecular biology. In: Hall, B. K.(ed) Homology: The hierarchical basis of comparative biology. Academic Press: San Diego, pp. 339–368.

Hodgkin, J./Horvitz, H. R./Jasny, B. J./Kimble, J. E. (1998) C. elegans: Sequence to biology. Science 282:2011.
Holst, E. von (1948) Vom Flug der Tiere und vom Menschenflug der Zukunft. Schriften der Universität Heidelberg 3:95–112.
Holst, E. von (1957) Wie flog Rhamphorhynchus? Natur und Volk 87:81–87.
Jolly, A. (1966) Lemur social behavior and primate intelligence. Science 153:501–506.
Katz, L. C./Shatz, C. J. (1998) Synaptic activity and the construction of cortical circuits. Science 274:1133–1138.
Lorenz, K. (1937) Über den Begriff der Instinkthandlung. Folia biotheoretica Serie B, 2, Instinctus:17–50.
Losos, J. B. (1990) The evolution of form and function: Morphology and locomotor performance in West Indian Anolis lizards. Evolution 44:1189–1203.
MacPhail, E. (1982) Brain and intelligence in vertebrates. Clarendon: Oxford.
MacPhail, E. (1985) Vertebrate intelligence: The null hypothesis. In: Weiskrantz, L. (ed) Animal Intelligence. Clarendon Press: Oxford, pp. 37–51.
Maddox, J. (1993) Has nature overwhelmed nurture? Nature 366:107.
Marler, P. (1991) The instinct to learn. In: Carey, S./Gelman, R. (eds) The epigenesis of mind. Lawrence Erlbaum Associates: Hillsdale NJ, pp. 37–66.
Maynard Smith, J. (1989) Evolutionary genetics. Oxford University Press: Oxford.
Mayr, E. (1958) Behavior and systematics. In: Roe, A./Simpson, G. G. (eds) Behavior and Evolution. Yale University Press: New Haven, pp. 341–362.
McLennan, D. A./Brooks, D. R./McPhail, J. D. (1988) The benefits of communication between comparative ethology and phylogenetic systematics: A case study using gasterosteid fishes. Canadian Journal of Zoology 66:2177–2190.
Owen, R. (1843) Lectures on the comparative anatomy and physiology of the invertebrate animals. Longman, Brown, Green & Longmans: London.
Patterson, C. (ed) (1987) Molecules and morphology in evolution: Conflict or compromise? Cambridge University Press: Cambridge.
Piaget, J. (1945) La formation du symbole chez l'enfant. Delachaux & Niestlé: Neuchâtel.
Plotkin, H. (1994) The nature of knowledge: Concerning adaptations, instinct and the evolution of intelligence. Penguin: London.
Prum, R. O. (1990) Phylogenetic analysis of the evolution of display behavior in the Neotropical manakins (Aves: Pipiridae). Ethology 84:202–231.
Reik, W./Surani, M. A. (eds) (1997) Frontiers in molecular biology. Oxford University Press: Oxford.
Rescorla, R. A./Wagner, A. R. (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A. H./Prokasy, W. F. (eds) Classical conditioning II: Current research and theory. Appleton: New York, pp. 64–99.
Riedl, R. (1977) Order in living organisms: Systemic conditions of evolution. Wiley: New York.
Ruvkun, G./Hobert, O. (1998) The taxonomy of developmental control in Caenorhabditis elegans. Science 282:2033–2041.
Science (1998) The First Gene Marker for IQ? Science 280(5364): 681.
Shettleworth, S. (1998) Cognition, evolution, and behavior. Oxford University Press: Oxford.
Simpson, G. G. (1958) Behavior and evolution. In: Roe, A./Simpson, G. G. (eds) Behavior and Evolution. Yale University Press: New Haven, pp. 507–535.
Skuse, D. N./James, R. S./Bishop, D. V. M./Coppin, B./Dalton, P./Aamodt-Leeper, G./Bacarese-Hamilton, M./Creswell, C./McGurk, R./Jacobs, P. A. (1997) Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. Nature 387:705–708.
Sulston, J. E./Schierenberg, E./White, J. G./Thomson, J. N. (1983) The embryonic cell lineage of the nematode Caenorhabditis elegans. Developmental Biology 100:64–119.
Takahashi, J. S./Pinto, L. H./Vitaterna, M. H. (1994) Forward and reverse genetic approaches to behavior in the mouse. Science 264:1724–1733.
Thomas, J. H. (1990) Genetic analysis of defecation in Caenorhabditis elegans. Genetics 124:855–872.
Thomas, J. H. (1994) The mind of a worm. Science 264:1698–1699.
Tinbergen, N. (1942) An objectivistic study of the innate behaviour of animals. Bibliotheca biotheoretica 1:39–98.
Tinbergen, N. (1951) The study of instinct. Oxford University Press: Oxford.
Tinbergen, N. (1959) Comparative studies of the behaviour of gulls (Laridae): A progress report. Behaviour 15:1–70.
Troemel, E. R./Chou, J.H./Dwyer, N.D./Colbert, H. A./Bargmann, C. (1995) Divergent seven transmembrane receptors are candidate chemosensory receptors in C. elegans. Cell 83:207–218.
Tully, T./Quinn, W. G. (1985) Classical conditioning and retention in normal and mutant Drosophila melanogaster. Journal of Comparative Physiology 157:263–277.
Weismann, A. (1892) Das Keimplasma. Eine Theorie der Vererbung. G. Fischer: Jena.
Wen, J. Y./Kumar, N. N./Morrison, G./Rambaldini, G./Runciman, S./Rousseau, J./van der Kooy, D. (1997) Mutations that prevent associative learning in C. elegans. Behavioral Neuroscience 111(2):354–368.

Gérard Weisbuch

The Complex Adaptive Systems Approach to Biology

Abstract

The purpose of this paper is to describe concepts and methods inspired from statistical physics of disordered systems and non linear physics and their application to theoretical biology. The central perspective is the study of functional organization of multi-component systems, based on a simplified description of individual components. The first section discusses a few examples of complex systems in physics and biology. We then describe three basic formalisms used in theoretical biology. The most important concept of attractor is introduced in the section on networks. We will discuss generic organization properties and the difference between organized and chaotic regimes. We will then propose two possible implementations of memory in the nervous and immune systems as examples of functional organization.

Key words

Complex systems dynamics, theoretical biology, neurons, immunology, Boolean nets, genetic regulatory network.

1. From Statistical Physics to Complex Systems

Statistical physics has accustomed us to mathematical descriptions of systems with a large number of components. The thermodynamic properties of ideal gases were understood as early as the end of the 19th century, while those of solids were understood at the beginning of the 20th century. In both cases, two important properties make modeling easy:

■ These are systems in which all of the components are identical.

■ If the interactions between the components are very weak, they can be ignored, as in the case of ideal gases. Otherwise, as in the case of solids, we can use linearization methods to put the problem into a form in which these simplifications can be made.

These early successes compared to the difficulties encountered in the understanding of biological systems would make us consider the above mentioned systems as rather simple. On the other hand, here are some examples of complex living systems:

■ The human brain is composed of approximately ten billion cells, called neurons. These cells interact by means of electrico–chemical signals through their synapses. Even though there may not be very many different types of neurons, they differ in the structure of their connections.

■ The immune system is also composed of approximately ten billion cells, called lymphocytes with a very large number of specificities which interact via molecular recognition, in the same way as recognition of foreign antigens.

■ Even the metabolism of a single cell is the result of the expression of a large number of genes and of the interactions among the gene products.

Although complexity is now a somewhat overused expression, it has a precise meaning within this text: *a complex system is a system composed of a large number of different interacting elements*.

In fact, the great majority of natural or artificial systems are of a complex nature, and scientists choose more often than not to work on systems simplified to a minimum number of components, which allows him or her to observe "pure" effects. The complex systems approach, on the other hand, is to simplify as much as possible the components of a system, so as to take into account their large number. This idea has emerged from a recent trend in research known by physicists as *the physics of disordered systems*.

1.1 Disordered systems

A large class of physical systems, known as multiphase systems, are disordered at the macroscopic level, but some are disordered even at the microscopic level. Glasses, for example, differ from crystals in that interatomic bonds in a glass are not dis-

tributed according to symmetries which we observe in crystals. In spite of this disorder, the macroscopic physical properties of a glass of a given composition are generally the same for different samples, as for crystals. In other words, microscopic disorder in a system does not lead to random global behavior. The simple models used by physicists are based on periodic networks, or grids, and simplified components of two different types are placed on the nodes, such as for example conductors or insulators in the problem known as percolation. These components are randomly distributed, and the interactions are limited to pairs of neighboring nodes. For large enough networks, we perceive that certain interesting properties do not depend on the particular sample created by a random selection, but of the parameters of this selection. In the case of the aforementioned insulator/conductor mixture, the conductivity between the two edges of the sample depends only on the ratio of the number of conductive sites to the number of insulating sites.

The percolation formalism exemplifies the approach taken by a number of theoretical biologists:

■ We choose to oversimplify the components of the system whose global behavior we would like to model. The formal genes, neurons and lymphocytes discussed below are cartoon-like simplifications of biological polymers and cells.

■ Nonetheless, these simplifications enable us to apply rigorous methods and to obtain exact results.

■ This approach is a dynamical approach. As in the differential methods, we start from a *local description* of the system, in terms of the short term state changes of the components as a result of their interactions. We expect the *global description* of the system from the method, that is to say the long term behavior of the system as a whole. The global behavior can be very complex, and it can be interpreted in terms of *emergent properties*. Within this notion is the idea that the properties are not *a priori* predictable from the structure of the local interactions, and that they are of biological functional significance (WEISBUCH 1990).

2. Networks

2.1 Units

Boolean automata. A simplified automaton is defined by its sets of inputs and outputs and by the *transition function*, which gives the output at time $t + 1$ as a function of the inputs and sometimes also the internal state (i.e., the output) at time t.

Boolean automata operate on binary variables, that is to say variables which take the values 0 or 1. In logical terms, 0 and 1 correspond to FALSE and TRUE, respectively. The usual logic functions AND, OR, and XOR are examples of transition functions of Boolean automata with two inputs. A Boolean automaton with k inputs, or of *connectivity* k, is defined by a truth table which gives the output state for each one of the 2^k possible inputs. There are 2^{2^k} different truth tables, and then 2^{2^k} automata.

Let $k = 2$. Here are the truth tables of four Boolean logic functions with two inputs:

	AND	OR	XOR	EQU
Input	00 01 10 11	00 01 10 11	00 01 10 11	00 01 10 11
Output	0 : 0 : 0 : 1	0 : 1 : 1 : 1	0 : 1 : 1 : 1	1 : 0 : 0 : 1

On the input line of the table, we have represented the four possible input states by 00, 01, 10, and 11. The four truth tables correspond to the standard definitions of the following logic functions: AND returns a 1 only if its two inputs are 1; OR returns a 1 only if at least one of its inputs is a 1; XOR is 1 only if exactly one of its inputs is a 1; and EQU the complement of XOR returns 1 when both input are equal. In logical terms, if A and B are two propositions, the proposition (A AND B) is true only if A and B are true.

We will further discuss the application of Boolean units to genetics.

Threshold automata. The state x_i of the ith threshold automaton is computed according to:

$$h_i = \sum_j J_{ij} x_j , \qquad (2.1)$$

$$x_i = 1 \text{ if } h_i > \theta_i \text{ ; } x_i = 0 \text{ otherwise}$$

The sum is computed over all of the inputs, subscripted by j. J_{ij} is the weight of the interaction between the ith and jth automata. In other words, the ith automaton has the value 1 if the weighted sum of the states of the input automata $\sum J_{ij} x_j$ is greater than or equal to the threshold, and 0 otherwise. Threshold automata are Boolean, but not all Boolean automata are threshold automata. We will further summarize some applications of threshold units to cognition (HERTZ/KROGH/PALMER 1990).

Formal lymphocytes. Not all networks are made of automata. A number of authors studying neural nets used differential equations as units. In immu-

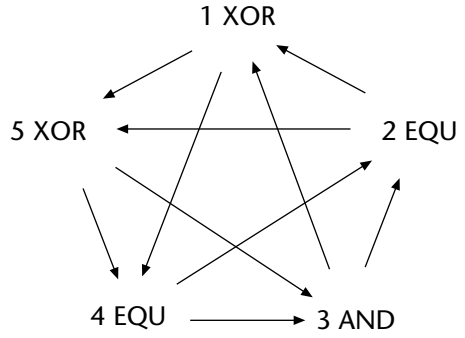

Figure 1: A network of five Boolean automata with two inputs. Each automaton has two inputs and transmits its output signal to two other automata. The XOR and AND functions have been previously defined. The EQU(ivalence) function is the complement of the XOR function—it is 0 only if exactly one input is a 1.

nology, PERELSON/WEISBUCH (1997) started from the following model, called the *B model* since it deals with B cells dynamics. The time evolution of the population x_i of clone *i* is described by the following differential equation:

$$\frac{dx_i}{dt} = m + x_i(pf(h_i) - d), \qquad (2.2)$$

where *m* is a source term corresponding to newly generated cells coming into the system from the bone marrow, the function $pf(h_i)$ defines the rate of cell proliferation as a function of the "field" h_i, and *d* specifies the per capita rate of cell death. For each clone *i*, the total amount of stimulation h_i is considered to be a linear combination of the populations of other interacting clones *j*. This linear combination is called the field, h_i, acting on clone x_i, i.e.,

$$h_i = \sum_j J_{ij} x_j \qquad (2.3)$$

where J_{ij} specifies the interaction strength (or affinity) between clones x_i and x_j. The choice of a *J* matrix defines the topology of the network. Typically J_{ij} values are chosen as 0 and 1. The most crucial feature of this model is the shape of the activation function $f(h_i)$, which is taken to be a log bell-shaped dose-response function

$$f(h_i) = \frac{h_i}{\theta_1 + h_i}\left(1 - \frac{h_i}{\theta_2 + h_i}\right) = \frac{h_i}{\theta_1 + h_i}\frac{\theta_2}{\theta_2 + h_i}, \qquad (2.4)$$

with parameters θ_1 and θ_2 chosen such that $\theta_1 << \theta_2$. Below the maximum of $f(h_i)$, increasing h_i increases $f(h_i)$; we call this the *stimulatory regime*. Above the maximum, increasing h_i decreases $f(h_i)$; we call this the *suppressive regime*. When plotted as a function of log h_i, the graph of $f(h_i)$ is a bell-shaped curve.

2.2 Structural Properties

An *network* is composed of a set of units interconnected such that the outputs of some are the inputs of others. It is therefore a directed graph, where the nodes are the units and the edges are the connections from the output of one unit to the input of another. Figure 1 represents the graph of the connections of a network of five Boolean automata with two inputs. This graph is equivalent to a set of five logical relations:

$$e(1) = XOR\ (e(2), e(3))$$
$$e(2) = EQU\ (e(3), e(4))$$
$$e(3) = AND\ (e(4), e(5))$$
$$e(4) = EQU\ (e(5), e(1))$$
$$e(5) = XOR\ (e(1), e(2))$$

where $e(i)$ is the state of the *i*th automaton.

2.3 Dynamical properties

Iteration mode. The dynamics of an automata network are completely defined by its connection graph (which automaton is connected to which), the transition functions of the automata, and by the choice of an *iteration mode*: It must be stated whether the automata change their state simultaneously or sequentially, and in what order. In the parallel mode, for instance, all of the automata change their state simultaneously as a function of the states of the input automata in the previous time step. Conversely, in the case of *sequential iteration*, or iteration in series, only one automaton at a time changes its state. Sequential iteration is therefore defined by the order in which the automata are to be updated. In the discussion that follows, we will talk only of *parallel iteration*.

Iteration graph. There are 2^N possible configurations for a network of *N* Boolean automata. The network goes from one configuration to the next by applying the state change rule to each automaton. Its dynamics can be represented by a directed graph, the *iteration graph*, where the nodes are the configurations of the network and the directed edges indicate the direction of the transitions of the network from its configuration at time *t* to a new configuration at time *t* + 1. Figure 2 represents

the iteration graph of the previous network (Figure 1) for the case of parallel iteration. This graph contains the $2^5 = 32$ possible states. It illustrates the fundamental dynamical characteristics which we will define below.

Attractors. Since an automata network is a deterministic system, if the network reaches a state for the second time, it will go through the same sequence of states after the second time as it did after the first time. Therefore, the system will go into an infinite loop in state space. These loops are called the *attractors* of the dynamical system, and the time it takes to go around the loop is called the *period* of the attractor. If this period is 1, as is the case for the configuration numbered 8 in the example, the attractor is a *fixed point*. We speak of *a limit cycle* if the period is greater than 1. The set of configurations which converge toward an attractor constitutes a *basin of attraction*. The network shown in the example below has four attractors.

Clearly it is only possible to construct a complete iteration graph for small networks. For the large networks we must be content to describe the dynamics of the system by characterizing its attractors. In this way we can try to determine:

- The number of different attractors,
- Their periods,
- The sizes of the basins of attraction (the number of configurations which converge toward each attractor),
- The notion of *distance* is also very important. The *Hamming distance* between any two configurations is the number of automata which are in different states.

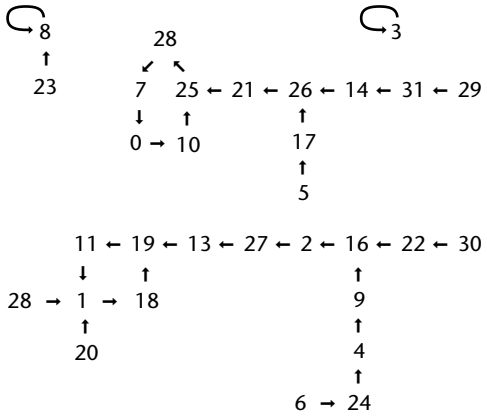

Figure 2: Iteration graph of the network of Figure 1. The numbers from 0 to 31 refer to the decimal representations of the 32 binary configurations of the network. The arrows show the temporal order of the configurations. Note that there are four different basins of attraction. State number 3 is an isolated fixed point. State number 8 is another fixed point. The other, larger, basins are composed of the configurations which converge toward the limit cycles with periods 4 and 5.

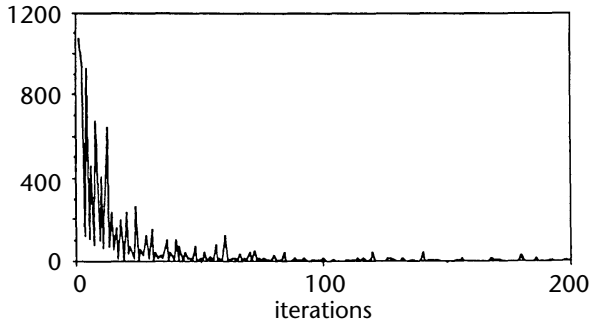

Figure 3: Histogram of the periods for 10 initial conditions of 1000 random Boolean networks of 256 automata.

3. In Search of Generic Properties

In view of all the simplifications that were made to define the units of the model networks, one cannot expect all properties of living systems to be modeled. Only some very general properties, independent of the details of the model will show-up. These are the so-called *generic* properties of the network. In fact, we are interested not in the particularities of a specific network, but in the orders of magnitude which we expect to observe in studying a set of networks with fixed construction principles. We therefore consider a set containing a large but finite number of networks. We choose some of these networks at random, construct them, and measure their dynamical properties. We then take the average of these properties, and we examine those which are fairly evenly distributed over the set of networks. An example will help to clarify these ideas.

Consider the Boolean networks with connectivity $k = 2$, with a random connection structure. The dynamical variable we are interested in is the period, for the set of all initial conditions and networks. Of course, this period varies from one network to the next. We have measured it for 10 randomly chosen initial conditions for 1000 different networks of 256 randomly connected automata, whose state change functions were generated at random at each node of the network. Figure 3 shows the histogram of the measured periods. This histogram reveals that *the order of magnitude of the period is ten* (this is the generic property), even though the distribution of the periods is quite large.

We can certainly construct special "extreme" networks for which the period cannot be observed before a million iterations. For this, we need only take networks which contain a random mixture of exclusive OR and EQUivalence functions (EQU is the complementary function of XOR; its output is 1 only if its two inputs are equal). But these extreme cases are observed only for a tiny fraction ($1/7^{256}$) of the set under consideration. We consider them to be pathological cases, *i.e.,* not representative of the set being studied.

We then call *generic properties* of a set of networks those properties which are independent of the detailed structure of the network—they are characteristic of almost all of the networks of the set. This notion then applies to randomly constructed networks. The generic properties can be shown not to hold for a few pathological cases which represent a proportion of the set which quickly approaches 0 as the size of the network is increased. In general the generic properties are either:

- Qualitative properties with probabilities of being true that are close to 1; or
- Semi-qualitative properties, such as the scaling laws which relate the dynamical properties to the number of automata.

The notion of generic properties characteristic of randomly constructed networks is the basis for the theoretical biological models. It has been extensively developed by physicists of disordered systems for the study of random microscopic systems such as glasses, or macroscopic multiphase systems. Physicists discovered (or rediscovered) many new concepts during the 70s. The notion of generic properties is similar to the notion of universality classes, developed for phase transitions. Without going into too much detail, we can say that the physical variables involved in phase transitions obey scaling laws which can be independent of the transition under consideration (such as, for example, phase transitions in magnetism, superconductivity, or physical chemistry) and of the details of the mathematical model which was chosen. These laws only depend on the physical dimension of the space in which the transition takes place (for us, this is three-dimensional space) and on the dimension of the order parameter. The set of phase transitions (and their mathematical models) which obey the same scaling laws constitutes a universality class. In fact, the first attempt to model a biological system by a disordered network of automata by S. KAUFFMAN (1969, 1993), a theoretical biologist, predates the interest of physicists in this subject. It is also based on the idea that the properties of disordered systems are representative of the vast majority of systems defined by a common average structure.

3.1 An example: Cell differentiation and random Boolean automata

The apparent paradox of cell differentiation is the following:

"Since all cells contain the same genetic information, how can there exist cells of different types within a single multicellular organism?"

Indeed, our body contains cells with very different morphologies and biological functions: neurons, liver cells, red blood cells… a total of more than 200 different cell types. Yet the chromosomes, which carry the genetic information, are not different in different cells. Part of the answer is that not all of the proteins coded for by the genome are expressed (synthesized with a non-zero concentration) in a cell of a given type. Hemoglobin is found only in red blood cells, neurotransmitters and their receptors only appear in neurons, etc.

Several mechanisms can interfere with the different stages of gene expression to facilitate or block it. We speak of activation and repression. The best known mechanisms involve the first steps of transcription. In order to transcribe the DNA, a specific protein, DNA polymerase, must be able to bind to a region of the chain, called the promoter region, which precedes the coded part of the macromolecule. Now, this promoter can be partially covered by a control protein, called the repressor; reading downstream gene is then impossible. It follows that, depending on the quantity of repressor present, the gene is either expressed or not expressed. The protein which acts as a repressor is also coded for by another gene, which is itself under the control of one or several proteins. It is tempting to model the network of these interdependent interactions by an automata network.

- A gene is then represented by an automaton whose binary state indicates whether or not it is expressed. If the gene is in state 1, it is expressed and the protein is present in large concentrations in the cell. It is therefore liable to control the expression of other genes.
- The action of control proteins on this gene is represented by a Boolean function whose inputs are the genes which code for the proteins controlling its expression.
- The genome itself is represented by a network of Boolean automata which represents the interactions between the genes.

In such a network, the only configurations which remain after several iteration cycles are the attractors of the dynamics, which are fixed points or limit cycles. These configurations can be interpreted in terms of cell types: a configuration corresponds to the presence of certain proteins, and consequently to the biological function of a cell and its morphology. Consequently, *if* we know the set of control mechanisms of each of the genes of an organism, we can predict the cell types. In fact, this is never the case, even for the simplest organisms. Without knowing the complete diagram of the interactions, S. KAUFFMAN (1969) set out to uncover the generic properties common to all genomes by representing them by random Boolean networks. Since there is a finite number of possible Boolean laws for an automaton with a given input connectivity k, it is possible to construct a random network with a given connectivity.

S. KAUFFMAN determined the scaling laws relating the average period of the limit cycles and the number of different limit cycles to N, the number of automata in the network. For a connectivity of 2, these two quantities seem to depend on the square root of N (although the fluctuations are very large). In fact, these same scaling laws have been observed for the time between cell divisions and for the number of cell types as a function of the number of genes per cell.

It is clear that KAUFFMAN's approximations were extremely crude compared to the biological reality—binary variables representing protein concentrations, Boolean (and thus discrete) functions, simultaneity of the transitions of automata, random structures… The robustness of the results obtained with respect to the possible modifications of the model (these are random networks) justifies this approach. As for the existence of a large number of attractors, it is certainly not related to the particular specifications of the chosen networks; it is a generic property of complex systems, which appears as soon as frustrations exist in the network of the interactions between the elements.

Presently, with the availability of many expression patterns, transcriptomes available thanks to DNA chips (BROWN/BOTSTEIN 1999), theoretists are facing a new challenge: how to deduce the network of gene expression regulation from the observation of the transcriptomes.

3.2 Generic properties of random Boolean nets

In fact, the results obtained by KAUFFMAN show two distinct dynamical regimes, depending on the connectivity. For networks of connectivity 2, the average period is proportional to the square root of N, the number of automata. The same is true of the number of attractors. In other words, among the 2^N configurations which are *a priori* possible for the network, the dynamics selects only a small number of the order of N which are really accessible to the system after the transient period. This selection can be interpreted to be an *organization* property of the network. As the connectivity is increased, the period increases much faster with the number of automata; as soon as the connectivity reaches 3, the period as well as the number of attractors become exponential in the number of automata. These periods, which are very large as soon as the number of automata is greater than one hundred, are no longer observable, and are reminiscent of the chaotic behavior of continuous aperiodic systems. In contrast with the organized regime, the space of accessible states remains large, even in the limit of long times. Further research (see DERRIDA 1987) has shown that other dynamical properties of these discrete systems resemble those of continuous chaotic systems, and so we will refer to the behavior characterized by long periods as *chaotic*.

Functional structuring. We have shown that when Boolean automata are randomly displayed on a grid their temporal organization in period is related to a spatial organization in isolated islands of oscillating automata as soon as the attractor is reached. In the organized regime, percolating structures of stable units isolate the oscillating islands. In the chaotic regime the inverse is true: few stable units are isolated by a percolating set of oscillating units (WEISBUCH 1990).

The phase transition. The connectivity parameter is an integer. It is interesting to introduce a continuous parameter in order to study the transition between the two regimes: the organized regime for short periods, and the chaotic regime corresponding to long periods. B. DERRIDA and D. STAUFFER (1986) suggested the study of square networks of Boolean automata with four inputs. The continuous parameter p is the probability that the output of the automaton is 1 for a given input configuration. In other words, the networks are constructed as follows. We determine the truth table of each automaton by a random choice of outputs, with a probability p of the outputs being 1. If $p = 0$, all of the automata are invariant and all of the outputs are 0; if $p = 1$, all of the automata are invariant and all of the outputs are 1. Of course the interesting

values of p are the intermediate values. If $p = 0.5$, the random process described above evenly distributes all of the Boolean functions with four inputs; we therefore expect the chaotic behavior predicted by KAUFFMAN. On the other hand, for values of p near zero, we expect a few automata to oscillate between attractive configurations composed mainly of 0's, corresponding to an organized behavior. Somewhere between these extreme behaviors, there must be a change of regimes. The critical value of p is 0.28. For smaller values, we observe small periods proportional to a power of the number of automata in the network. For $p > 0.28$, the period grows exponentially with the number of automata.

Distances. The distance method has recently been found to be one of the most fruitful techniques for determining the dynamics of a network. Recall that the Hamming distance between two configurations is the number of automata in different states. This distance is zero if the two configurations are identical, and equal to the number of automata if the configurations are complementary. We obtain the relative distance by dividing the Hamming distance by the number of automata.

The idea of the distance method is the following: we choose two initial conditions separated by a certain distance, and we follow the evolution in time of this distance. The quantity most often studied is the average of the asymptotic distance, measured in the limit as time goes to infinity. We compute this average over a large number of networks and of initial conditions, for a fixed initial distance. Depending on the initial distance, the two configurations can either evolve toward the same fixed point (in which case the distance goes to zero), or toward two different attractors, or they could even stay a fixed distance apart (in the case of a single periodic attractor), regardless of whether the period is long or short. Again, we observe a difference in the behaviors of the two regimes. On Figure 4, obtained with a cellular connectivity (the network is a regular two-dimensional grid), the x-axis is the average of the relative distances between the initial configurations, and the y-axis is the average of the relative distances in the limit as time goes to infinity. In the chaotic regime, we observe that if the initial distance is different from 0, the final distance is greater than 10%. The final distance seems almost independent of the initial distance. On the other hand, in the organized regime, the final distance is proportional to the initial distance.

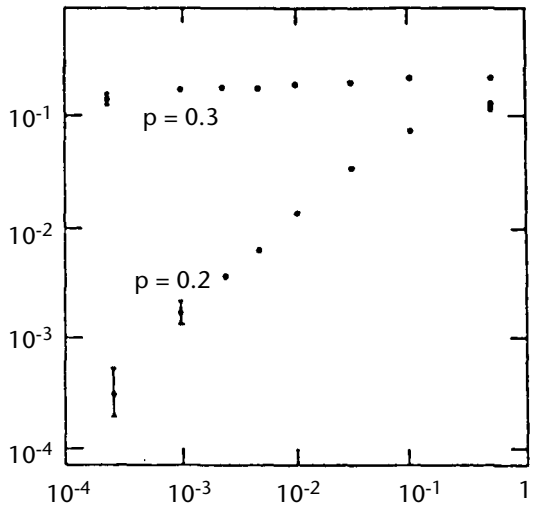

Figure 4: Relative distances at long times as a function of the initial relative distances, in the organized ($p = 0.2$) and chaotic ($p = 0.3$) regimes (from DERRIDA/STAUFFER 1986).

Property	Organized regime	Chaotic regime
Period	small	large
Scaling law (periods)	goes as a root of N	exponential in N
Oscillating nodes	isolated subnetworks	percolate
Distance	d_∞ proportional to d_0	d_∞ remains finite

Table 1: Generic properties of random networks differ according to the dynamical regime, organized or chaotic. N is the number of automata in the network, d_0 is the distance (always taken small with respect to N) between two initial configurations, and d_∞ the distance between the two evolved configurations at large times.

Conclusions. This study clearly demonstrates the existence of two types of behaviors, organized and chaotic. Table 1 summarizes the differences in the generic properties of these two regimes.

4. Memories

4.1 Neural nets and distributed memories

There now exists a very large literature on neural nets which we are not going to report here (HERTZ/KROGH/PALMER 1990). Let simply summarize the results. Neural nets with symmetrical connections have an exponential number of point attractors. This result applies to random serial iteration, and

exponential means that the logarithm of number of attractors is proportional to the number of units.

Neural nets are most often used in learning tasks. A general learning algorithm is HEBB's rule. When reference patterns (network configurations) are presented to a network to be learned, connections can be constructed that ensure that the attractors of the network dynamics are the reference patterns. Furthermore the dynamics drives the network from initial conditions not to far from the reference patterns to the nearest reference patterns: these nets can then be used as associative memories that can be recalled from partial memories.

HEBB's rule can be written:

$$J_{ij} = \sum_{\mu} S_i^{\mu} S_j^{\mu}$$

where μ refers to the different reference patterns and S_i and S_j to the states of connected neurons i and j in the corresponding pattern.

Memories are thus distributed in the network as opposed to a memory that would be localized on some part of the net. The memory capacity of a fully connected neural net build according to HEBB's rule scales as the number of units in the net: no more than 0.14 N patterns, where N is the number of units, can be stored and retrieved in a HOPFIELD neural net.

4.2 Immune nets and localized memories

As a memory device, the immune system needs to obey certain constraints (PERELSON/WEISBUCH 1997): it should be sensitive enough to change attractor under the influence of antigen. It should not be too sensitive and over react when antigen is present at very low doses. The immune system should also discriminate between self-antigens and foreign antigens. Finally, it should be robust—memories of previously presented antigens should not be lost when a new antigen is presented. Thus, in some sense, the system should be able to generate independent responses to many different antigens. This independence property is achieved when attractors are localized, i.e., when the perturbation induced by an encounter with an antigen remains localized among the clones that are close to those that actually recognize the antigen (see Figure 5).

Our problem is to classify the different attractors of the network and to interpret the transitions from one attractor to another under the influence of antigen perturbation.

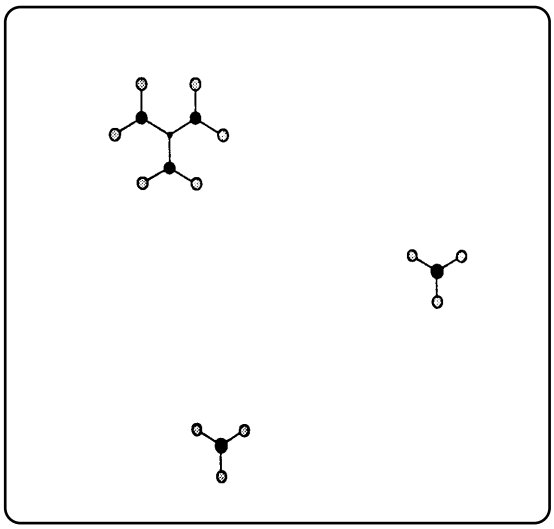

Figure 5: Localized patches of clones perturbed by different antigenic presentations. Two vaccination and one tolerant attractors are represented.

Let us start with the most simple virgin configuration, corresponding to the hypothetical case where no antigen has yet been encountered and all populations are at level m/d, i.e., all proliferation functions are 0. After presentation of the first antigen, memorization is obtained if some populations of the network reach stable populations different from m/d. In the case of a localized response, there will be a patch close to the antigen specific clone in which cells are excited out of the virgin state. Each antigen presented to the network will result in a patch of clones that are modified by the presentation. As long as the patches corresponding to different clones do not overlap, the various antigens presented to the network can all be remembered. Once the idea of localized non-interacting attractors is accepted, everything is simplified: instead of solving 10^8 equations, we only have to solve a small set of equations for those neighboring clones with large populations, supposing that those further clones that do not belong to the set have populations m/d. A practical approach to studying localized attractors is to combine computer simulations and analytic checks of the attractors by solving the field equations (see below).

Immunity. Let us examine the case of antigen presented to clone Ab_1, which results in excitation of

clones AB_2, clones AB_3 remaining close to their virgin level (see Figure 6). We expect that AB_1 will experience a low field, L, while AB_2 will experience a large suppressive field, H. From the field equations we can compute the populations x_i. Recall, from Eqs. (2.2) to (2.4),

$$h_1 = zx_2 = L = \frac{d\theta_1}{p'} \tag{4.1}$$

$$h_2 = x_1 + (z-1)\frac{m}{d} = H = \frac{p'\theta_2}{d} \tag{4.2}$$

where $p' = p - d$.

An immune attractor is usually reached for an intermediate initial antigen concentration, and intermediate decay constants, If the initial antigen concentration is too low or if the antigen decays too fast, the immune attractor is not attained and the system returns to the virgin configuration, i.e., AB_1 and AB_2 populations increase only transiently and ultimately return to the virgin m/d level. Thus, no memory of antigen encounter is retained.

Tolerance. Another localized attractor corresponds to tolerance (see Figure 7).

A strong suppressive field acts on AB_1 due to AB_2's, the AB_2's proliferate due to a low field provided by AB_3's, but AB_4's remain nearly virgin. The field equations once more allow one to compute the populations:

$$h_2 = x_1 + (z-1)x_3 = L = \frac{d\theta_1}{p'} \tag{4.3}$$

which gives x_3 if one neglects x_1, which is small.

$$h_3 = x_2 + \frac{(z-1)m}{d} = H = \frac{p'\theta_2}{d}, \tag{4.4}$$

and thus for small m/d

$$h_1 = zx_2 \approx zH \tag{4.5}$$

Substituting h_1 in Eq. (2.2) gives a very small value for $f(h_1)$, which shows that x_1 is of the order of m/d. The AB_1 population, experiencing a field several times higher than H, is said to be *over-suppressed*.

As in the case of the immune attractor, one can study the conditions under which the tolerant attractor is reached when antigen is presented. One finds that tolerance is obtained for large initial antigen concentrations, slow antigen decay rates and large connectivity, z (PERELSON/WEISBUCH 1997).

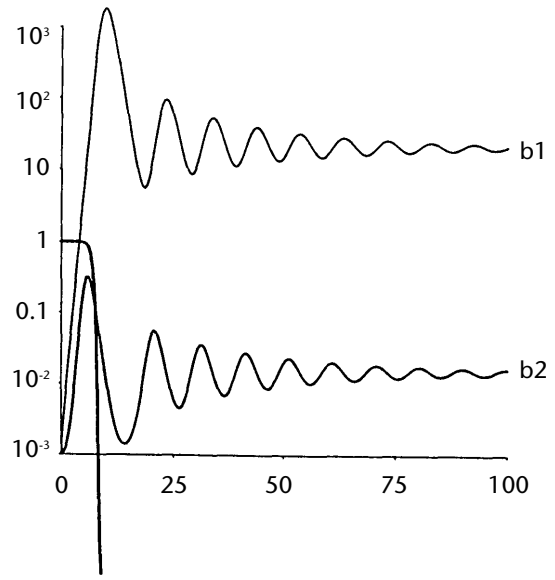

Figure 6: Time plot of an antigen presentation resulting in a vaccination attractor. On the vertical axis are the clone populations on a logarithmic scale. Time in days is on the horizontal axis. In the vaccinated configuration the largest population is localized at the first level. X_1 is high (H) and sustained by an intermediate population (L/z) of X_2. The rest of the clones are virgin (V) (or almost virgin) after the system settles into this attractor. When antigen is presented again, it is eliminated faster than the first time.

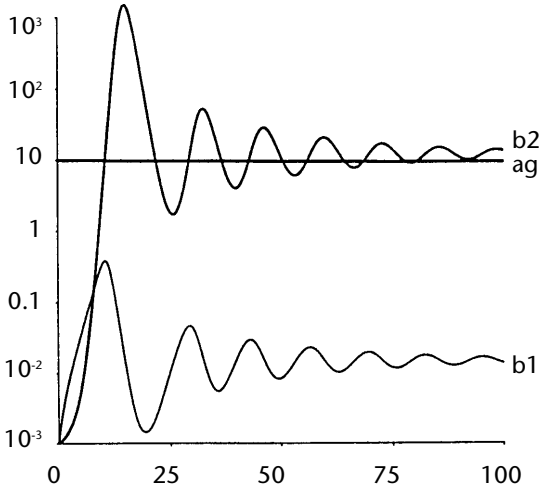

Figure 7: Time plot of an antigen presentation resulting in a tolerant attractor. X_2 is high (H) and sustained by an intermediate population (L/z) of X_3. X_1 is over-suppressed by the X_2 and is not able to remove the antigen.

4.3 Number of attractors

Localized attractors can be interpreted in terms of immunity or tolerance. Because these attractors are localized they are somehow independent: starting from a fully virgin configuration, one can imagine successive antigen encounters that leave footprints on the network by creating non-virgin patches, each of these involving a set of p perturbed neighboring clones. An immune patch contains $1 + z$ clones, a tolerant patch $1 + z^2$ (see Figure 5). Independence of localized attractors implies a maximum number of attractor configurations that scales exponentially with N, the total number of clones. The following simplified argument gives a lower bound. Divide the network into $N/(1 + z^2)$ spots. Each spot can be in 3 possible configurations: virgin, immune or tolerant. This gives a number of attractors that scales as $3N/(1 + z^2)$. Few of these attractors are of interest. The relevant question is the following: A living system must face frequent encounters with antigen during its life. Self antigen should elicit a tolerant response; dangerous external antigens should elicit immune responses and subsequent immunity. The nature of the localized response on each individual site of the network is then determined by the fact that the presented antigen should be tolerated or fought against. In this context, we can ask how many different antigens can be presented so that no overlap among different patches occurs.

In the case of random antigen presentation, simple reasoning (WEISBUCH 1990; WEISBUCH/OPREA 1994) is sufficient to derive the scaling law relating m, the memory capacity (i.e., the maximum number of remembered antigens) to N, the total number of clones. Let n_s be the number of suppressed clones involved in a patch.

m is given by:

$$m \cong \sqrt{\frac{2N}{n_s}}$$

and this provides an estimate for the mean memory capacity of the network.

The only assumption to obtain this scaling law is the random character of the network with respect to antigens, i.e., the network is not organized to respond to the set of presented antigens. On the other hand, it can be argued that the clones expressed by mammals have been selected by evolution according to the environment of the immune system, e.g., to be tolerant to self molecules and responsive to frequently encountered parasites and pathogens. If the system were optimized to the antigens in its environment, the network could be filled compactly with non-overlapping patches. The number of antigens (patches) would then scale linearly, i.e.,

$$m \propto \frac{N}{n_s}.$$

WEISBUCH/OPREA (1994) discuss more thoroughly the capacity limits of model immune networks with localized responses. They verify by numerical simulations the square root scaling law for the memory capacity. They also examine a number of other features of the network. They show that when the number of presented antigens increases, failures to remove the antigen occur since the relevant clone has been suppressed by a previous antigen presentation. They also show that previous immune or tolerant attractors are rather robust in the sense that destruction of these local attractors by new encounters with antigen is rare, and that the complete re-shuffling of the attractors, as in HOPFIELD nets (HERTZ/KROGH/PALMER 1990), is never observed.

5. Conclusions

This presentation does not mention a number of interesting topics such as the origin of Life and Species, issues in population ecology, metabolic networks, etc. The selected topics that I developed reflect only my own research interests. Since the first version of the paper was written new experimental techniques and the availability of huge series of data changes the outlook for theoretical biology.

- A number of empirical studies concerned the topology of interactions and pointed to the importance of scale free networks, i.e., networks with a wide distribution of connectivities, such as food-webs, gene regulatory nets, metabolic networks… (ALBERT/BARABASI 2002).
- The availability of huge data sets allow to envision the solution of inverse problems: how to compute the set of interactions from the set of observed network configurations.
- But the importance of characterizing the generic properties of empirical or simulated networks still remains a preliminary step before further studies.

Author's address

Gérard Weisbuch, Laboratoire de Physique Statistique, Ecole Normale Superieure, 24 rue Lhomond, F-75231 Paris Cedex 05, France. Email: weisbuch@lps.ens.fr

References

Albert, R./Barabasi, A. L. (2002) Statistical mechanics of complex networks. Reviews of Modern Physics 74:47. lanl archive: cond-mat/0106096.

Brown, P. O./Botstein D. (1999) Exploring the New World of the genome with DNA microarrays. Nature Genetics 21:33–37.

Derrida, B. (1987) Dynamical phase transitions in random networks of automata. In: Souletie, J./Vannimenus, J./Stora, R. (eds) Chance and Matter. North Holland: Amsterdam.

Derrida, B. /Stauffer, D. (1986) Phase transitions in two-dimensional Kauffman cell automata. Europhysics Letters 2:739.

Hertz, J. A./Krogh, A. /Palmer, R. (1990) Introduction to the theory of neural computation. Santa-Fe Institute Studies in the Sciences of Complexity. Addison-Wesley: Redwood City CA.

Kauffman, S. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. Journal of Theoretical Biology 22:437–467.

Kauffman, S. (1993) The origins of order: Self-organization and selection in evolution, Oxford University Press: New York.

Perelson, A./Weisbuch, G. (1997) Immunology for physicists. Reviews of Modern Physics 69:1219–1267.

Weisbuch, G. (1990) Complex systems dynamics. Santa-Fe Institute Studies in the Sciences of Complexity. Addison-Wesley: Redwood City CA.

Weisbuch, G./Oprea, M. (1994) Capacity of a model immune network. Bulletin of Mathematical Biology 56:899–921.

Carlos Stegmann

The Human Behavior Instinct: How Decisions for Action Are Reached

An Interdisciplinary Inquiry into the Nature of Human Behavior

Abstract

The main tenet of the proposed approach is that human behavior is controlled by a universal behavior program*, explaining how the* information *concerning behavior-governing rules originates. There exist two fields of possible cognition: objective cognition or heteronomy, related to objects or purposes, and moral cognition or autonomy, related to moral norms of conduct. Phylogenetic development has created in humans two proximal mechanisms for cognition, which are adapted to the two fields, and their operation is based on the process of "imprinting" discovered by Konrad* LORENZ *and Niko* TINBERGEN*. What gets to be imprinted are* beliefs *in the truth or rightness of objective and moral norms, which govern decisions for action. In the case of objective cognition the imprinting is reversible, whereas moral cognition operates with irreversibly imprinted beliefs upheld with "ontomoral rigidity" (*KANT*'s "categorical imperative"). This rigidity bases the evolution of the moral norms that sustain the "extended order" of human society (*HAYEK*). The beliefs are affixed to emotions which we perceive as action-inducing feelings (*DAMASIO*).*

Key words

Behavioral biology, human behavior, evolution, morals, ethics, cognition, interdisciplinary analysis, epistemology.

"MY PROBLEM IS… THAT WE need not only an evolutionary theory of knowledge but also an evolutionary theory of morals, that this evolutionary theory of morals is only now beginning to be conceived, and that its essential feature will be that morals are not a creation of reason, but a second tradition independent from the tradition of reason, which helps us to adapt to problems which exceed by far the limits of our capacity of rational perception" (Friedrich A. v. HAYEK, inaugural address to the Alpbach Forum 1985).

1. Introduction

The main tenet of the attempted explanation of human behavior is that this behavior is controlled by an innate, phylogenetically developed instinct. The present paper describes the main elements of this attempt. Based on these elements, the author's book "Evolution of Morals" contains discussions and interpretations of works of some of the main philosophers and scientists who have devoted themselves to the subject. The manuscript of this book has been deposited at the Konrad Lorenz Institute for Evolution and Cognition in Altenberg, Austria.

Biological explanations of human behavior have so far been based on extensions of the findings of behavioral research applied to non-human living organisms, such as the attempts made by sociobiology and human ethology. These attempts have provided deep insights into the human behavior apparatus, confirming in every way that humans possess a complex and variegated "parliament of instincts" (the expression coined by Konrad LORENZ), as do our non-human relatives of the animal kingdom. The reduction of group behavior to individual behavior was mainly accomplished by William HAMILTON through his thesis of "kin selection", presented in his seminal paper of 1964, "The Genetical Evolution of Social Behavior", from which the concept of "inclusive fitness" was derived. This thesis was taken up by Edward WILSON, who, in combining it with the principles of population genetics, created the new science of Sociobiology (WILSON 1975). In his application of this theory to the human species, WILSON (1978) was able to show, without leaving any doubt, that human beings carry phylogenetically developed "propensities" (i.e., in-

stincts), including "…bonding between parents and children, heightened altruism toward closest kin, incest avoidance, suspicion of strangers, tribalism, …" and many other traits. He then goes on, however, to say "…people have free will and the choice to turn in many directions…". In "On Human Nature" (1978, p6) he poses the question:

"Human emotional responses and the more general ethical practices based on them have been programmed to a substantial degree by natural selection over thousands of generations… Which of the censors and motivators should be obeyed and which ones might better be curtailed or sublimated?"

The circularity of this question, which has been noted neither by WILSON nor by all the authors who put the same question in various other forms, is the central problem posed by attempts to explain human behavior. The attempt to extend the theories of sociobiology to the human species collapses as it reaches the formidable obstacle of human "free will". What is needed is a theory to explain the *human decision process*.

WILSON (1978, p208) also calls those "censors and motivators", a "jerrybuilt foundation of partly obsolete Ice-Age adaptations", maintaining, as many of his colleagues do (TOOBY and COSMIDES among other, as quoted in HESCHL 1998[1]), that human behavioral instincts were formed by genetic evolution as adaptations to conditions prevailing in the Pleistocene, and are totally unable to cope with the requirements of the modern world. To be able to cope, as Konrad LORENZ said (1992), "[the] rigid instinctual ways of behavior [require] the surveillance by our rational, responsible morals".

The notion that human behavior is governed by "rigid instincts" has been bitterly rejected by authors like S. ROSE, R. C. LEWONTIN and L. J. KAMIN (1984), who consider that to defend such a notion amounts to the (for them) repugnant moral attitude of "biological determinism". Here another problem appears, which affects the discussion of human behavior: the apparently unavoidable intromission of moral attitudes[2] into the discussion. Ethics, a philosophical discipline, has governed this discussion in recorded history, and it is quite difficult to open the subject to objective thinking.

Both the attitude of "biological determinism" and its rejection, implying a revulsion against the idea of a dependence on natural laws, are based on the mistaken logic that natural biological laws, like physical laws, must necessarily imply automaticity, robbing humans of their free will. But natural laws only answer the "how" question, their formulation provides tentative (synthetic) explanations of observed phenomena; and if the phenomena are human attitudes and behavior which are anything but automatic, the laws will have to explain how the *changes* of the behavior-governing rules and norms come about. It is the attitudes themselves, both WILSON's "curtailing and sublimating" and ROSE et al.'s censure, and including also LORENZ's "rational, responsible morals", which must be explained by an objective theory of human behavior. "Freedom", "free will", are *analytic* concepts, which bear no relation at all, and cannot be taken, by themselves, as having a *synthetic* correspondence to anything existing in reality. The real problem for a theory of human behavior is the disclosure of how the *information* about the norms and values, which govern this behavior, originates.

This status is claimed by the present proposed explanation. It submits that humans possess a phylogenetically developed instinctual apparatus, which operates, like the similar apparati of our non-human relatives, through the generation of action-commanding feelings. But the feelings generated by the human behavior instinct are not elicited automatically as are the action-motivating emotions of our non-human relatives, but are linked to objectively known values and rules. As such, the human behavior instinct is overimposed, integrates, and controls the "parliament of instincts" of our genetic heritage. The mechanism is still an instinct, but there is this difference which makes us human: we *know* what we are doing.

This specifically human behavior instinct can be equated to a *universal behavioral program*, in analogy to the *universal grammar program,* which governs the learning of language (PINKER 1995). Just as words and their meaning, parsing rules and other culturally evolved characteristics of language are "imprinted" in the receptive grammar program through the agency of group members, so are culturally evolved norms of behavior "imprinted" in the receptive behavior program, also through the agency of group members. The way in which this happens, the various involved mechanisms and processes, are the subject of the proposed explanation.

Human beings are not limited to biology. To be able to produce credible results, an attempt to explain something as pervadingly inclusive as human behavior, must include all the manifestations of this behavior, from the cognitive to the physiological. Moral cognitive phenomena in *homo sapiens* occur simultaneously as cultural phenomena in large social groups, as behavioral phenomena in the individ-

uals who compose these groups, and as neurophysiological phenomena in the neural systems of the same individuals. The natural laws which govern the phenomena at the sociocultural level of groups are the subject of political, social, and economic science; phenomena of social behavior coincide and are simultaneous with behavioral phenomena in the individuals who compose the groups, governed by laws which are the subject of behavioral biology and psychology; in turn, behavioral phenomena in individuals coincide and are simultaneous with the generation, at the neurophysiological level, of conceptual representations associated with emotions and feelings, which command decision making and action. The approach of the present inquiry thus is to attempt a thoroughgoing inter-level, i.e., interdisciplinary reduction of human behavior, from political science and sociology to behavioral biology and psychology to neurophysiology. It is based on, and is consistent with, the findings of many researchers and scientists, but mainly on the works of Immanuel KANT, Friedrich HAYEK, Konrad LORENZ, Sigmund FREUD, Rupert RIEDL, Judith HARRIS, Adolf HESCHL, and Antonio DAMASIO.

In spite of its repeatedly checked and demonstrated internal and external consistency, the present account is still only an *explanation in principle,* and is mainly intended to be a basis for research programs of the involved disciplines, to conduct the tests and statistical surveys that may confirm or falsify it. One prediction is made, however, in relation to Antonio DAMASIO's "somatic markers' (DAMASIO 1994), which may be verified by neuropsychological tests (cf. Section 9).

2. Postulates, Basic Concepts, and Terminology

Postulates originate from the faculty of the human mind to imagine logical extremes which do not occur elsewhere in the reality of nature, but are a precondition for the cognition of reality, or better: they are the actual *means* of cognition. Their formulation originates in the *analytic* capacity of the human mind, which is used to produce *synthetic* statements about reality (KANT 1904, CARNAP 1995). This means that the logical condition whereby a postulate is non-demonstrable, or self-evident, or has an ontological nature, is irrelevant for theories that aim to represent accurately facts of this world. Postulates are structural parts of those theories and must be justified by the success of their application (VOLLMER 1990).

In the following, some of the postulates and basic concepts on which the theory is based are stated. Also, in an attempt to avoid misunderstandings as much as possible, in the Section "Terminology" some of the more relevant terms that have been used in the context of the theory are defined.

Postulates and basic concepts

Evolution is the basic impelling force of life since that moment in the dawn of our living world when commands for growth and reproduction were first encoded in a genome, including behavior consistent with these goals. Thus "purpose" became a part of nature. Living beings cannot do otherwise: growth and reproduction is a necessity. Evolution of life causes decrease of entropy and increase of order in an open system. The object of this evolution is information. Since something *new* cannot be known beforehand, because if it could, it would not be new, the only possible mechanism of *creating* information is the random mutation-and-selection mechanism of evolution, the genome being its sole repository. Any other kind of explanation would have to resort to supra-natural causes (HESCHL 1998). All *existing* information is thus innate, it can only be put to use if *already possessed* by the genomes of particular individuals,—otherwise they would never be able to develop their cognitive/behavioral abilities. Although the *origin* of *all* information resides in the genome, the cognitive tasks are *performed* through the proximate cognitive "apparati", the seat of which are those vast brain and nerve systems whose structural and operational design, incorporated in the genome, has been created by the evolutionary mutation-and-selection mechanism.

According to the philosophers of the Age of Reason, humans would possess only one kind of intellectual apparatus, embodied by the term "reason", a premiss still prevalent in our time. KANT (1904, p19):

"[…it is required from] a critique of practical reason that, once completed, it must constitute a unity with speculative reason, representable through a single common principle, because in the end there can exist only one and the same reason, differing only in its application".

Although more than 200 years have elapsed since KANT formulated this thought, it still constitutes an implicit basic assumption, or premiss, in modern theories of mind and behavior. It leads to the false conceptualization that the process which determines the formation of moral information would be a process of objective cognition governed

by reason. It will be seen from the following discussion that such a process is not possible. The attempted explanation contends, instead, that human beings possess *two* distinct cognitive apparati which, although related to each other, operate in different ways. One is the apparatus called *reason,* which is the seat of objective cognition ("speculative reason") as understood by KANT. The other is the apparatus called *conscience*, the seat of moral cognition or (also following KANT) practical reason. In actual fact, KANT's observation must be considered to be correct in the sense that we obviously do have in our heads a single and very complex apparatus of cognition. What is postulated is that cognition operates according to two very different *processes* when trying to develop *information* related to an object or purpose in the field of *objective cognition* or *heteronomy*, and information related to general laws of conduct in the field of *moral cognition* or *autonomy*.

A second basic assumption of these philosophers which is contradicted by the theory is that reason and its faculties would belong to an "intellectual" world distinct and independent from the "natural" biological world. The proposition that there exists an "intellectual world" separate from nature, or a dichotomy nature–intellect, is senseless, and so are all those dichotomies like matter–spirit, nature–culture, instinct–reason, and many similar ones which can all be subsumed in the "body–soul" dogma, the absurdity of which has been conclusively demonstrated by Gilbert RYLE (1949). To assume the existence of any of these dichotomies means to commit a fundamental error of category. Matter and spirit do not exist separately but represent different strata of the integrated entity homo. The entity homo, to whom the phenomena of spirit belong, is a new, integrated category of existence, and not a machine inhabited by a ghost, as RYLE, quite crassly, and, as he said, "with deliberate abusiveness", put it. Spiritual events such as morals do not pertain to some esoterical realm separate and independent from nature, but must be considered to be natural phenomena; they are stratum manifestations of events that occur simultaneously in all the strata of the integrated entity. They are thus explainable by natural laws pertaining to their stratum but also, as a special form of explanation, through their reduction to the laws of the lower strata that compose the integrated entity. Nevertheless, the assumption that there exists a purely intellectual or cultural realm independent from nature, where speculative and moral reasoning takes place and philosophers pursue the discipline of ethics, is still implicit in all the sociobiological and ethological writings which have been consulted.

The process, which determines the formation of moral information, takes place at the level, or stratum, of individual human behavioral physiology. This stratum, the seat of the human behavioral apparatus "conscience", is interposed between the stratum of social and cultural phenomena and the stratum of neurophysiological phenomena. The idea of strata was taken from Konrad LORENZ' "levels of integration" within "systemic wholes" (LORENZ 1973, pp56ff), where new levels of integration appear as a result of the phenomenon of "fulguration", a concept developed by LORENZ (1973, pp47–50) which is related to Gilbert RYLE's quoted idea of category. Living beings, including homo sapiens, are integrated natural units or entities (systemic wholes) composed of strata, from the atomic-particle to the socio-cultural strata. Stratum-specific natural laws apply in each stratum, but events or phenomena which affect the entity are one-and-the-same event or phenomenon in all and each of the strata. Since the event or phenomenon is one-and-the-same, there cannot be any "causation" between strata, but the natural laws which apply in each stratum must be reducible to the laws which apply in the hierarchically lower strata; the idea of natural law only makes sense in this context. Another consequence is that all the strata of an integrated unit evolve together; there can be no strata with differing rates of evolution, as of society and of the human beings which compose it—the evolution of both is one-and-the-same phenomenon. Since genetic evolution is a process much too slow to account for the evolution of human behavioral physiology simultaneously with the rapid evolution of human social order, there must be another mechanism at work, consistent with the faculty of the human species of transmitting and receiving information to and from group members.

The categories of moral behavior are those described by Immanuel KANT in his work "Elements for a Metaphysics of Morals" ("Grundlegung zur Metaphysik der Sitten") (KANT 1904). KANT's fundamental insight into human behavior has not changed in the more than 200 years since it was formulated. What appears to be new is the explanation of moral behavior in terms of natural laws, which KANT could not give, limited, as he was, by the scientific advances of his times. The present theory can thus be regarded as an attempt to provide a physical-science interpretation of KANT's categories of moral behavior.

Terminology

The theory is composed of words which establish relationships between several special words considered to be the main terms of the theory: truth and rightness; to believe and belief; morals and moral behavior, ethics, and values; cognition, knowledge and information; and objective and moral cognition. Following POPPER (1976), the definitions which are given below do not purport to establish any more profound meaning of the defined terms, independently, that is, dissociated from the context of the theory.[3] The definitions are strictly ad hoc or functional, i.e., the meaning attributed to the terms is strictly and exclusively what is meant by them in the context of the theory. (Common definitions are, unless noted, from American Heritage Dictionary.)

In a field as controversial as the discussion of human behavior it is important to try to avoid misunderstandings, which are usually centered on the meaning of words. I thus appeal to readers to consider the given definitions, and not others, in their analyses of this inquiry.

Truth, rightness. In the context of the theory, *truth* is defined by the correspondence concept. Thus, *truth* is simply "*the correspondence of a statement with the facts that it intends to describe*". Truth is one of those logical extremes, or forms of cognition, which the analytic faculty of the human mind is capable of imagining, but which do not occur elsewhere in nature. Thus truth is forever unattainable, forever preliminary. For as soon as a correspondence seems to be established it becomes subject to criticism, and, eventually, it will be superseded by other truths, as new facts and correspondences are discovered. If applied to a moral norm, value or standard held to be true, its *truth* is expressed as *rightness*.

To believe, belief. "*To believe*" is here meant in the transitive acceptation of this verb, as requiring an object to complete its meaning: "*to uphold or maintain, to believe in the truth of a statement*".

Beliefs, a thoroughly human cognitive category, govern human behavior. *Beliefs*, in the context of the present theory, are *emotional attachments to objective concepts or mind images, which uphold the truth* (q.v.) *of those concepts or images*.

Morals, moral behavior, ethic(s), and values. The term "*morals*" designates, in the context of the theory, the kind of behavior which distinguishes the human species. Thus "*morals*" are "*norms and institutions which govern human behavior*" without any reference to the values good, right, wrong, bad or evil. Obviously all norms and institutions, qua norms and institutions, are supported by values considered to be good, or exclude others considered to be bad. The theory itself, however, is objective: it takes into account that such values exist without supporting or condemning any of them, nor does it intend to establish any kind of method whereby what is good or bad may be ascertained. The theory thus envisages the same attitude of objectivity, of scientific detachment, applied in behavioral studies of irrational beings. *Moral behavior* is *behavior related to universal instrumental and end values (norms and institutions)*, it being indifferent, or not relevant (from an objective point of view), if this behavior follows, or not, established values, or some new values which are different from the established ones.

The term "*ethics*" is reserved for "*moral cognition and information*", i.e., cognitive processes to establish, and information about, values like good, right, wrong, bad or evil. Ethical information resides, as will (it is hoped) become clear in the exposition, in ethical beliefs.

Strictu sensu, "*moral values*" are only those already mentioned: good, right, bad, wrong, or evil. However, in the course of cultural evolution, norms and institutions have always been associated with those values, and were thus also referred-to as "*values*". As such they have been classified into "*instrumental values*" (*norms*) and "*end values*" (*institutions*) (MOHR 1987, pp8–9).

Cognition, knowledge, and information. The term *cognition* is used only in one of its two acceptations: "*the mental process or faculty by which knowledge (information) is acquired*", or, "*the process of recognizing information about an object*". The other acceptation, "that which comes to be known", is not intended when the term is being used in the present inquiry.

In relation to *knowledge*, dictionaries, as usual, cannot evade the circularity implied by the use of some synonym, which in this case is the word whose meaning is closest to the meaning of "*to know*", which is "to apprehend". The definition of "to know", given in most dictionaries, is "to apprehend (to know) with certainty", and the term *knowledge*, as used in the context of the theory, is "*that which is known with certainty*". The term, however, is anthropomorphous, it refers to a content of the human mind, and the qualifying expression "with certainty" must be taken cum grano salis, as it is always based on a *belief* (q.v.). To avoid as much as possible

the non-relevant connotations, which the human-centered meaning of the word "knowledge" can imply, it becomes appropriate to use instead the well-established physical magnitude *information* (cf. Section 7).

Objective and moral cognition. With the appearance of life there also began the division of the world into a knowing "subject", and a knowable external "object", represented by a certain environment (HESCHL 1998). In the most common acceptation an "object" thus always presupposes a "subject", and "objective" is the opposite of "subjective". In the present inquiry, however, "objective" is always used in conjunction with "cognition" (q.v.), and the acceptation of "objective" is the common "*uninfluenced by emotion, surmise, or personal prejudice*", but is here qualified more stringently as *implying compliance with the Principle of Objectivity*, which says that, *to arrive at valid results and conclusions, the process of objective cognition must be separated from, and totally indifferent in relation to moral values*. The "object" of objective cognition is what is called "reality", "the real world", or "nature", and thus, ultimately, objective cognition pursues the discovery of natural laws. Moral values, that is, the rules and standards which govern human behavior, as an object which exists in nature, are also subject to objective cognition in a quest to discover the natural laws which govern the process of their formation and evolution. This process has here been called *moral cognition*. It is *the process of recognizing information about the rules and standards which govern human behavior*, and is, as such, an object (a "subject") of *objective cognition*. The present inquiry is an *objective* attempt to discover the natural laws and mechanisms which govern *moral cognition*.

3. Basic Hypothesis: The Imprinting of Beliefs

As seen in the previous section, spiritual events such as morals do not pertain to some esoterical realm separate and independent from nature, but must be considered to be natural phenomena, explainable by natural laws. It seems obvious that moral phenomena *have* to be reduced to that apparatus we call "conscience". There is an enormous literature on it, and to say that human conduct is governed by conscience is a hackneyed commonplace. Strange as it seems, nevertheless, this may be considered some sort of discovery, as the term conscience is almost never mentioned in ethological texts about human behavior and its comparison with that of other species.

The phylogenetic evolution which culminated with the appearance of homo sapiens must have included several instances of the phenomenon "fulguration", whereby, through the combination of various elements, a new category of existence comes into being, the properties of which could not have been predicted from the properties of the composing elements (LORENZ 1973, pp47–50). The new entity now possesses a new upper stratum, but it is still the integrated whole of all the strata that compose it, from the atomic-molecular to the spiritual. Certain natural laws act in every stratum, and explanations or reductions must be considered possible. The new species is distinguished, chiefly, by the possession of a new cognitive apparatus, which permits the transmission of learned or acquired information and thus initiated a new form of evolution. (OESER 1983, 1996, 1997)[4]. This apparatus operates with two different mechanisms or processes:

- The mechanism or process of *objective cognition*, usually called *reason*, which contains the *analytic* faculty which permits to derive *synthetic* knowledge about reality (CARNAP 1995)[5], combined with the faculty of *language* and a greatly expanded *memory*, and which superseded (but still includes), as a result of the fulguration, the *ratiomorph faculties* of the pre-human ancestor;
- The mechanism or process of *moral cognition,* usually called *conscience*, which relates objectively known *options for action* to, also objectively known, *values*, which in turn are related to certain *feelings* which the apparatus generates to make us act (cf. Section 9), and which superseded (but still includes), as a result of the fulguration, the *action-governing instincts* of the pre-human ancestor.

The present inquiry is an attempt to comprehend and explain the laws and processes which determine the operation of the apparatus *conscience*. It does not pretend to make a similar attempt to discover the structure and way of operation of the apparatus *reason*, which is probably one of today's greatest challenges of science. It had to refer, however, to the relations and differences between moral and objective cognition, and this subject is treated below, supported by KANT's division of *pure (objective)* and *practical (moral)* cognition (KANT 1904, 1966).

"Conscience" is *the human behavior instinct*, the phylogenetically developed *universal behavioral program*, possessed identically by all humans. It operates by relating behavioral decisions and corresponding actions to objectively known norms and values, and

to the motivating emotions affixed to those norms and values. It acts, restraining or enhancing, upon every member of the "parliament of instincts" which our species inherited from its ancestors.

Since "instinct" is a recurring concept in the present inquiry, it is appropriate to give its definition, as understood by behavioral biology. From Rupert RIEDL, "Biology of Knowledge" (1979, p289):

"'Instinct', or species-specific instinctual action is understood as a hereditarily predetermined action, the drive of which is totally endogenous (without external stimulus). The instinctual action is set in motion by an innate release mechanism (IRM) and will always run off in the same species-specific manner. If there occurs a retarding of the release the threshold release value will get steadily lower until the action will be released without an external reference object, resulting in the so-called idle motion. In many cases an instinct is related to a previous appetence behavior, a search for the biological reference object".

The instincts of our non-human relatives, as can be seen, are "hereditarily predetermined", and "will always run off in the same … manner". This is the "rigidity" or "automaticity" which the opposers of "biological determinism" complain about. The human behavior instinct, however, being anything but automatic, does not respond entirely to this definition. It is an "instinct" because, like the conventionally defined instincts, it is also a phylogenetically developed, hereditarily predetermined control of behavior. It will also always run off in the same human-specific manner. The difference is that it does not anymore correspond to the control of single, automatically released actions, but relates to all kinds of behavior, being overimposed on all those other instincts (still possessed by homo), which it controls—albeit quite often with increasing difficulty, as, for instance, when the threshold release value of an aggressive action cannot be restrained in the face of a received insult. This control and its relation with the innate release mechanism (IRM) are discussed below.

The question now arises as to the operating mechanism of this instinct. We all know that objectively known norms and values govern our moral behavioral decisions and corresponding actions. But these norms and values do not exist as genetical inheritance in our minds. How are they introduced?

In his work "The Backside of the Mirror' (1973, pp278–284), Konrad LORENZ has postulated a process which he called "the striving for renewal of youth". There he says that, "…there comes a time when the developing youth puts into critical question all the traditional values received from his or her parents, and seeks for new ideals". This phenomenon, according to LORENZ, is phylogenetically programmed for its survival value, as it ensures the continuous adaptation of norms and values to a changing external reality:

"The surprisingly quick process of attachment to a new cultural group, the fixation of the instinct of collective enthusiasm [zeal] to a new object, carries features which remind strongly of a process of object-fixation known from the animal kingdom, the so-called *imprinting*. As in this process, [the fixation] is linked to a certain sensitive development phase of youth, is independent of training factors, and it is irreversible, at least insofar as a first attachment of this kind can never be followed by a second one of the same intensity".

The phenomenon of "imprinting" was discovered by Konrad LORENZ and Niko TINBERGEN (for this and other features discovered through behavioral research they were awarded the 1973 Nobel Prize for medicine, LORENZ 1978) The "innate schemata" proposed by this theory to exist in the animal mind define features of conspecific objects likely to be "imprinted". The present theory posits that in humans the "imprinting" phenomenon refers to beliefs in the rightness of moral values, which are always associated with cultural attachments to groups with definite ideals and mores, that is, with very human conspecifics. Thus, the main step which led to the formulation of the explanation of human behavior proposed in this inquiry is the assertion that the fixation of norms and values in our behavior apparatus not only "reminds strongly", but that it *IS* the process of object-fixation known as *imprinting*. The basic hypothesis can thus be expressed as follows: *the operating mechanism of the human behavior instinct is based upon the process known as "imprinting" in behavioral biology, and the contents of the imprints are beliefs in the rightness of certain norms and values*.

From Rupert RIEDL, "Biology of Knowledge" (1979, p292):

"As imprinting we consider that special part of a learning process whereby the learning content can be assimilated only during a [certain short] phase of the biological development, and remains irreversibly engraved thereafter. Some organisms learn by i. the images of their parents or sexual partners. The basic physiologic mechanism is open for any learning content, thus facilitating experimental procedures with organisms which possess this faculty. The mechanism may likewise be extended to human beings, who may be said to be imprinted by the conditions of their civilization".

As rightly observed by RIEDL, the mechanism must also be present in that member of the animal kingdom, the human species. In the other species of that kingdom the process elicits a series of pre-programmed hereditary instinctual responses of behavior which are affixed upon the imprinted object. It is contended in the present hypothesis that in the human species the mechanism also elicits a series of behavioral responses, which *are affixed upon the imprinted beliefs in norms and values*. It is possible to compare this imprinting mechanism in the human species to the language learning process, which is also a process of imprinting, in this case of words, in the pre-existing open learning program of universal grammar (PINKER 1994). What get to be imprinted in the moral learning program of conscience are beliefs in values. But what actually happens in the imprinting process, is the affixation of endogenously generated *feelings*, to certain objectively known values; it is these feelings which govern the selection between options for action in the decision-making process (see Section 9). It may also be observed that, if humans, as maintained by WILSON or by TOOBY and COSMIDES, would possess *only* the rigid instincts formed in the Pliocene/Pleistocene period, they wouldn't really be able to *learn*. A "learning" based on those instincts, could only be equated to animal training. The proposed universal behavior program, instead, renders true learning possible, through imprinting, and all the other proposed mechanisms.

In non-human beings the instinctual behavior is activated by the Innate Release Mechanism (IRM). It can be surmised that in the human behavior instinct this mechanism has been superseded by the *decision-making process*. This may be illustrated by an example given by Konrad LORENZ. In "On Aggression" (1963, pp60–61 and 248–249) he refers to the so-called "polar sickness" to demonstrate the consequences of holding back aggressive urges if no normal release outlet is available. LORENZ himself lived one such situation in a prison camp during the war, and

"… if I did not hit my friend but kicked and trampled an empty carbide canister instead, my action was due entirely to my knowledge of the symptoms of instinct withholding… Our insight into the causal chains of our own behavior actually can give our reason and our morals the power to intervene in situations in which the categorical imperative, relying on itself alone, would fail miserably".

LORENZ attributes the cause of his behavior to his specialized knowledge about "the symptoms of instinct withholding". *The pertinent question to ask here, however, is: how did the feelings originate which motivated LORENZ to try every possible means to avoid hitting his friend*? Clearly, when LORENZ noticed that his aggressive urge was nearing its release value, *he reached a decision governed by a moral value*. The action released by this decision was to discharge the accumulated aggressive urge on a non-sentient thing instead of hitting his friend.

It may be noticed too that contrary to the alleged failure of the categorical imperative, it was this imperative, according to the physical-science interpretation given in the present account, which originated the feelings that governed the decision process which led to LORENZ' moral action.

It must be recognized that the example of LORENZ' behavior is anecdotal, and cannot be taken, individually, to base any theory. However, as stated in the Introduction, the present is mainly an *explanation in principle* proposed as a basis for research programs for the involved disciplines, and the present example is well suited to clarify a main issue involved in the theory. It may thus be stated that the present explanation of *the human behavior instinct* is essentially an explanation of *the human decision process*.

4. Reversible and Irreversible Beliefs

Human actions follow objective rules to attain some aims or purposes. The selection of aims or purposes, and of certain constraints to do, or refrain from doing, imposed on behavior, however, are governed by moral end- and instrumental values. As seen, the objective rules, and the end- and instrumental values, correspond to two fundamentally different fields of possible cognition existing in reality, which determine the possibility of making predictions and of substantiating the rules and values with objective critical arguments. It must be supposed that the behavior-governing apparatus is adapted to these two fields, considering how different the human attitudes concerning recognition of information corresponding to the two fields are.

The field of objective cognition is KANT's *heteronomy*, the field where knowledge (information) results from the interaction of objects with human cognition. We apply the *apriori-analytic* faculty of our minds to derive *aposteriori-synthetic* statements, i.e., theories, about the objects (KANT 1904, 1966; CARNAP 1995; LORENZ 1983). A theory is a conjectural statement about facts related to the objects, composed of two parts: ascertainment and explanation, from which a prediction of new facts is derived which must be substantiated by tests or by observation; the truth content of a theory is thus its predictive capac-

ity. Once the tests or observations are made, confirming, or not, the predictions of the theory, a new set of facts is established which, in turn, will give rise to new theories. The process can be represented as an unending ascending spiral, the ascent representing the increase of information (knowledge) (RIEDL 1979; POPPER 1990). This is by now a well-known tenet of the evolutionary theory of knowledge, being representative of what is known as the "scientific method". The success of this method, which in modern times was instrumental in producing large advances of objective knowledge, led, however, to the erroneous conviction that it could be applied to all kinds of problems, including the design of the norms and institutions which base the order of human society. The error resides in assuming that predictions and their critique are as possible in the realm of moral cognition, the realm of the social order, as in the domain of objective cognition. To uncover the nature of this error it is necessary to look more closely at what predictions entail, and to examine the possibility *in principle* of making predictions in the fields of objective, and of moral cognition.

The controlling indicators of objective cognition are the predictions of new facts, relating either to the attainment of certain objectives or to new object-related facts. Predictions are thus the essential component of this field of knowledge, and, together with their critique, are always possible, as the problems to be solved are normally clearly put within well-defined boundaries set by the objectives to be achieved.

Physical reality concerning the possibility of making predictions is entirely different in the field of ethics, the realm of KANT's *autonomy*. Ethical information is composed of instrumental and end values, i.e., laws of behavior which are distinguished by their universal, purposeless character, and institutions, which, based on ethical decisions, are perceived as general aims. These are the laws and aims which base HAYEK's *extended order* of society (HAYEK 1982, 1988). Any change of these general laws and aims will provoke so many intricately interwoven direct, indirect, and retroactive effects, that their analysis would involve combinatorial numbers of astronomical dimensions, and thus make a prediction of their overall effects on the structure of the extended order all but impossible. Besides, even if such a computation were at all possible, it is by no means granted that the predicted new structure of the extended order would be considered "good" by contenders and critics alike. This is but another way of putting the old facts/standards ("Sein/Sollen") problem, embodied by the "naturalistic fallacy": facts cannot be inferred from standards, nor can standards be derived from facts. The only means to develop information in the moral field is the trial-and-error method of evolution. It is only by trying out a standard, norm or aim, that *ex-post-facto* its appropriateness or adaptive fitness can be ascertained; but no cogent prediction of this appropriateness or adaptive fitness can be made, exactly as in genetic evolution, which must contend with the same kind of physical reality. In the moral field the only possible testing field is human society itself, where a moral value, in view of the intransigence of both contenders and critics, can only be implemented through hard struggle and fight, the intransigence being a necessary requisite in that fight (cf. Section 7).

Based on these considerations it becomes possible to establish an evolutionary interpretation of the proposed belief fixation. In the field of heteronomy, actions are governed by what KANT called "problem-related assertory rules", enacted by his "hypothetical imperatives" of dexterity and wisdom. In this field, in which predictions of consequences, and soundly based critical arguments are always possible, the ease of *collaboration* between members of a group has a high survival value. It must be supposed, therefore, that the imprinting of beliefs corresponding to this field is basically "*reversible*", which is amply corroborated by everyday observation: although beliefs in the appropriateness of certain rules and methods to attain pre-established aims are held with executive firmness by the responsible individuals, they will agree to changes if well-argued critiques make this necessary.

In the second field, the field of morals proper, which comprises the selection of aims, and of general (purposeless) norms of behavior, actions are governed by what KANT called "apodictic" rules, enacted by his "categorical imperative" of morals. In this field, as seen, the physical or mathematical reality of the naturalistic fallacy implies the practical impossibility of making adequate predictions, and of submitting the selection to rationally argued criticism. The selection and establishment of adequate aims and rules is, however, of paramount importance for group survival (success). Considering the physical reality, the only means of establishing the appropriateness of aims and rules is to submit them to the trial-and-error process of evolution, for which it becomes necessary to uphold them with unyielding stiffness. The proposition of the theory is thus that through phylogenetic adaptation the imprinting of the beliefs of this field is basically *irreversible*, being

concomitant with attitudes and ways of behavior in support of the beliefs, like disposition to self-sacrifice, and willingness to fight (and kill). It must also be supposed that the categorical imperative of universality, and "the absolute compulsion of the will to follow certain rules of conduct with no express purpose" (KANT 1904, p49), represents a phylogenetically evolved category of cognition. The apparatus will recognize a belief as corresponding to this category, and activate the irreversible fixation, and the behavioral attitudes for its support. Knowledge (information) in this field is essential for the establishment of the *extended order* of human society by governing *competition* between individuals and social groups, and is carried, through a phylogenetic adaptation, by *irreversible beliefs*, whereby ethical standards, norms and aims are upheld with the utmost rigidity, as a necessary condition, given the impossibility of predicting their outcome, for their chances of getting implemented.

5. Formation of Beliefs in Individuals and Groups

The imprinting of beliefs occurs during the formative years of each individual, being a part of the postulated moral learning program. As in the documented similar learning programs of irrational beings, the "openness" is restricted to certain biologically pre-programmed periods of life. The process of learning and imprinting is socially determined, and the programs must be considered to have evolved by phylogenetic adaptation to the physical and social reality.

The engraving of reversible beliefs in the purpose-dependent heteronomic rules which govern the field of objective cognition seems to remain open during the whole life of an individual. The imprinting of the irreversible ethical beliefs, however, is genetically time-programmed, being "open" only during a sensitive period in the development of each individual (LORENZ 1973, p282). It appears to take place in two phases. In the first phase, which occurs from infancy through puberty, ethical values are impressed unto the learning youngsters, as their moral learning program will be open to receive the imprinting from their social environment, following certain natural laws. These laws have been ascertained by Judith HARRIS in her thesis of "group socialization" (HARRIS 1998). According to this thesis, since their earliest age children manifest a strong tendency to associate in peer groups, and accept and engrave certain manners and norms of conduct established in association with these peers, thus bonding them in a group; or, as is mostly the case, the youth's association will occur with some already existing group of peers, whose "culture" the newly arrived will adopt. The groups are possessed by a strong sense of solidarity for the defense of their distinguishing manners and values. This thesis contradicts the commonly held notion that the engraving of ethical beliefs is accomplished through the agency of the youth's parents ("The Nurture Assumption", the title of HARRIS' book). To the contrary, the engraving through peer groups usually represents a rebellion *against* the values held by parents—unless, as HARRIS recognizes, the values held by the peer groups are representative of larger social groups to which the parents also belong. HARRIS' thesis covers the formative period of young humans, from infancy to adolescence. But children, teenagers, and adolescents are still immature; they still can change their beliefs. At some moment, however, between adolescence and early adulthood (maturity), a *final identification* with a belief system will take place, which will mark the individual for life. As seen in Section 3, Konrad LORENZ (1973) has postulated this stage of the process. It is in this last stage when the final, irreversible imprinting of beliefs in instrumental and end values takes place. It must be recognized that it is sometimes possible for intelligent individuals to extricate themselves from a "wrong" belief, one that continuously clashes with reality. This is exemplified by Arthur KOESTLER (1967), whose realization of the clash of his Communist belief with reality and reason caused him to substitute for it the belief in the pathological nature of all beliefs (the "ghost" in the machine). Such a disengagement from an ethical belief in later years is not very common, and is accompanied by almost unsurmountable difficulties, as KOESTLER's example shows. But always, so it seems, the result will be the conversion of the individual to another ethical belief. For beliefs we must have. Reason is always fallible, whereas beliefs are firm and trustworthy, not subject to doubts. They are the strong, redoubtable guides which lead humanity's evolutionary progress, without which humanity could not go on.

The values, but most especially the end values, are the "causes" or "ideals" with which each individual identifies him or herself. They are provided by the social groups with which the young adult associates, participating in certain "movements", which are usually influenced by well-known social reformers, philosophers, religious or political leaders which may be considered as "codifiers" of the Zeitgeist of each historical epoch (FRIEDELL 1976[6]). *The individual*

engraving of beliefs is thus also the ground for solidary group behavior. It is, as stated in a postulate of Section 2, the same phenomenon or event taking place simultaneously in all the strata of a natural entity. The causation is external, consisting of changes of the physical, technical and economic reality, which require adaptations of the general rules of conduct, instrumental and end values, to ensure survival. Groups and individuals together, operating according to the proposed mechanism, will respond to the challenge. Together with other spiritual manifestations, the values and institutions compose the Zeitgeist of each historical epoch. But it is by no means assured that values and institutions defended by "progressive" or "conservative" groups and individuals will prove to be successful in coping with the external challenges. The process of selection of values is evolutionary, this being the subject of Section 7.

6. Fields of Cognition

The foregoing considerations permit now to establish the structure of a system comprising the two described fields of cognition and their relation to the field of human behavior. Human actions normally follow plans to attain some purpose. *Planning* is an activity which partakes of both fields of cognition, leading to *decision making* and to *action*. The ethical field governs the commitments to aims or purposes, often embodied by institutions, and also provides the restrictions and restraints to action carried by the irreversible beliefs in universal, purposeless rules and standards of conduct, enacted by KANT's categorical imperative. The field also supplies the statistical information required for planning[7].

The field of objective cognition supplies the information required for the design of the means and steps to attain the purposes, often in the form of applied science. The resulting purposeful rules of action are carried by the reversible beliefs enacted by KANT's hypothetical imperatives of wisdom and dexterity. Every plan involves a prediction, and since the executive rules of action originate in the field of objective knowledge, both such predictions and their critique are eminently possible, so long as the universal rules and aims originated in the field of ethical information are not changed.

The diagram of Figure 1 visualizes the main elements of both fields of cognition and their relation to the activity of planning. The equivalence imperatives/somatic markers is explained in Section 9, which contains a more detailed description of the process of *decision making*.

7. Order, Information and Evolution

Order and information

The preceding sections refer to what may be considered the fundamental part of the present inquiry, in that they postulate the basic mechanism whereby the controlling norms of human behavior are formed. The nature of this mechanism is evolutionary; that is, the human cognitive apparatus "conscience" produces *information* about the *validity* of moral norms by way of the trial-and-error method of evolution. Since these two concepts have various acceptations, it is necessary to explain their contextual meaning and application in the present theory, and their relation to evolution.

In the Section "Terminology" it is stated that the term "information" is used in the present inquiry in preference to "knowledge". The latter term cannot be defined precisely; its meaning is subjective and may vary in the view of the beholder, whereas "information" is a well-established physical magnitude. To consider its application in the present theory it is necessary to review the derivation of the relevant physical concept. In what follows this derivation is summed up following Erwin SCHRÖDINGER's "What is Life?" (1944) and Rupert RIEDL's extension of these concepts to the order of living beings, presented in his book "The Order of Life" (1975).

There is a general tendency of thermal decay in the Universe, measured, according to the 2d law of thermodynamics, by the increase of the thermal magnitude entropy. There is an intuitively evident link between this decay, measured by the increasing entropy, and increasing "disorder". This link was the subject of a profound study by Ludwig BOLTZMANN towards the end of last century, whereby he gave it a mathematical expression which establishes the equivalence of entropy and disorder. In this equivalence, "information" is both a measure of entropy and of disorder, that is, of the indetermination of events occurring in a real-world message-emitting source, or system. Its unit of measurement is the "bit" if it relates to disorder, or entropy units (cal/^{o}K) if it relates to entropy.

As shown by SCHRÖDINGER in his book, the main distinguishing mark of life is that living beings counteract that general tendency of decay. In their limited realm, which is but a small part of the universe (where entropy and disorder still increase), they create order and reduce entropy. They do this through the commands for growth and reproduction, that is, according to SCHRÖDINGER, the "plans of develop-

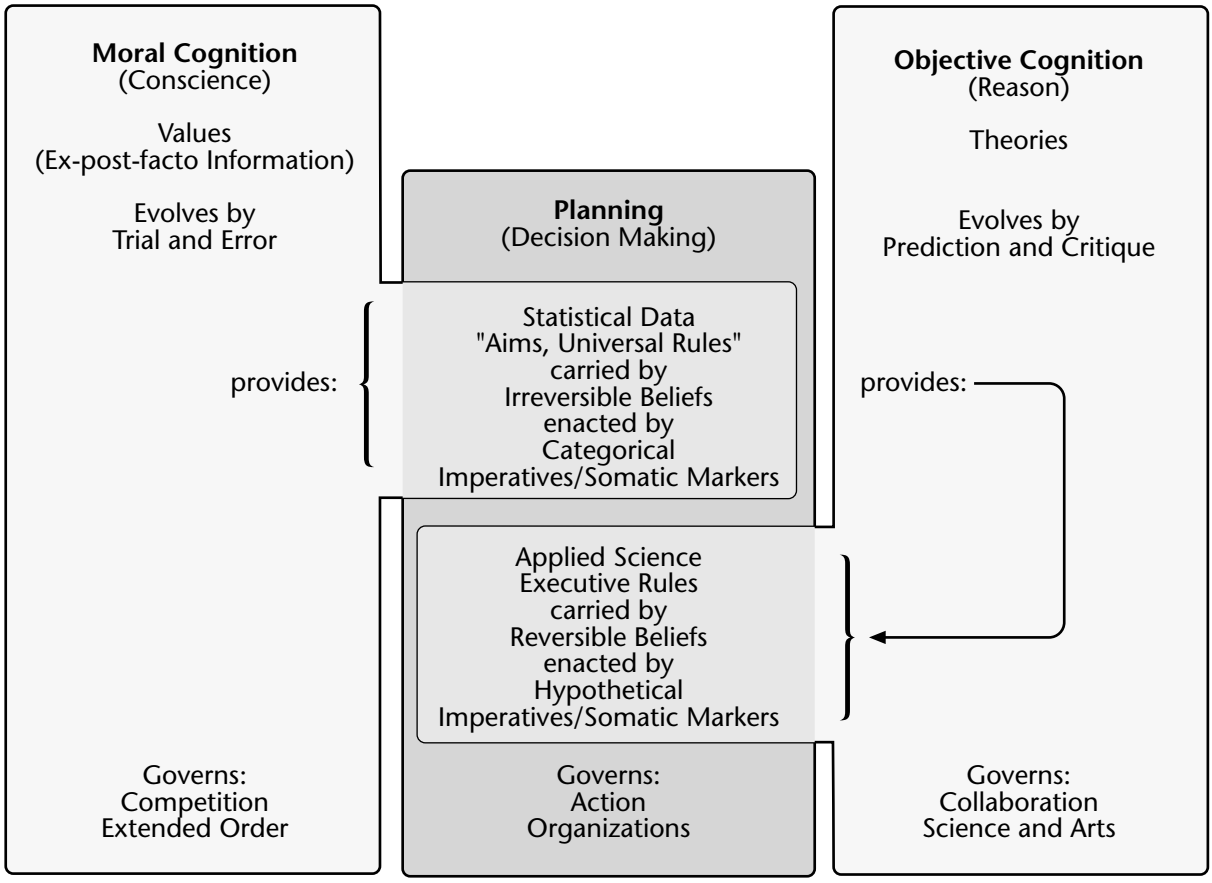

Figure 1: Fields of Cognition

ment" which are encoded in their genomes. Information can thus also be considered as a measure of "negative entropy" or "negentropy", and order, that is, of the determination of events in a real-world message-emitting source, or system. The units of measurement are the same as for entropy and disorder.

As seen above, with the advent of life there appeared order and the concept of information as determination content of living systems. RIEDL expresses this as the theorem "information equals order or negentropy". The order of living systems is determined by the information contained in their genomes. RIEDL calls this information "law-content", and order is the product of this law-content (expressed in bits) times the number of applications. And here the important notion of "validity" must be introduced. Order is created by *valid* information. RIEDL illustrates this through an example given by Lwoff, whereby, if in a certain gene a guanine molecule is exchanged for an adenine molecule, the information content remains the same from the point of view of the physicist; the mutation, however, may prove to be lethal, the organism dies, and the contribution to order is nil.

The commands for growth and reproduction encoded in the genomes of living organisms produce their multiplication according to the "plans of development" contained in the genetical information content. In this process the chance events of mutations occur, causing variations of the "plans of development". These will either prove unadapted and die off (most of them), or produce organisms with living ability of a different kind; this is the trial-and-error process characteristic of evolution, and the "order-on-order" principle formulated by SCHRÖDINGER ensures that it will always produce an increase of information, order, and negentropy. The *validity* of the order-producing information simply means that the organism which is built according to the structural genetic information, and follows the behavior prescribed by its genetically encoded instincts, has been successful *until now*; but no *prediction* can be made about its continued success, nor can predictions about the success or insuccess of mutated organisms be made. The principle of natural selection discov-

ered by DARWIN is not a natural law. It is simply the ascertainment of a fact; it says, essentially, "organisms which are fit are selected, and those which are not fit, are not selected", or, to put it more simply, "organisms will either be selected or they will not be selected". This kind of statement is called a tautology in logic—it is true in any case, and no predictions can be made based on that principle.

What is here contended is that all the concepts discussed above are applicable to the behavior of human beings. The information content of a moral norm is comparable with the information contained in the genetically encoded behavior-controlling instincts of non-rational animate organisms—comparable, that is, in relation to the considered conceptual aspects related to the theory of information, but by no means identical. In both cases, the information content will either succeed in establishing itself as a component of order, or "die off"—but, in relation to this, two main differences (among other) must be mentioned immediately:

■ In non-rational living organisms what succeeds or dies off, are the organisms themselves; in the case of human beings what succeeds or "dies off" are the *norms*;

■ In non-rational organisms the information content of the behavioral programs known as "instincts" (see definition in Section 3) is "teleonomic", it belongs to the realm of heteronomy, of purposeful behavior, within the limiting frame provided by the general laws of nature. In the case of human beings the information content of moral norms is "autonomic", it belongs to the realm of autonomy, as these norms are the general, purposeless laws which provide the limiting frame of the "extended order" of human society. This "extended order", an expression which we owe to Friedrich HAYEK, is superimposed on the real-world order governed by the general laws of nature. To the extent in which the behavior of human beings is heteronomic, that is, so far as it is intended to attain certain concrete purposes[8], it is also limited by the general laws of nature.

These moral norms, following the previously discussed physical concepts, are the *information* residing in the determination content of the order of human society, and the concept of "validity" applies to them in the same way as for the behavior-governing instincts of non-rational beings. Here, however, a third main difference has to be added to the two mentioned previously: instincts are followed *automatically* by the non-rational beings, whereas humans always have the possibility of contravening a moral norm. However, *we do know about the validity of norms*, even without being able to demonstrate this validity with our objective cognitive faculties, which, as has been seen, is impossible, being an expression of the "naturalistic fallacy". Gerhard VOLLMER (1996, pp79–80), for instance, is intent on demonstrating how, departing from certain basic norms other norms can be derived. As an example he mentions one "basic norm" (whose content is not relevant for the present argumentation) "which is recognized by practically everybody, … a general axiom which, although undemonstrable, is nevertheless convincing". VOLLMER and many other people are convinced of the validity of such a norm, and once it reaches this status, once it is "recognized by practically everybody", the norm will become a part of the information content of the extended order.

At this point one can return to the baffling idea of "*Free Will"* mentioned in the Introduction. "Free Will" is one of these abstract categories which exist only in the analytic framework of our minds but tell us *nothing* about reality. Being present in our mind, however, leads (deludes) our reason to consider them as synthetic conceptual characterizations of facts of the real world, which they aren't. The idea of "Free Will" *contributes nothing to the disclosure of objective information about the natural mechanisms which regulate human behavior.* The real problem at issue is, as seen, the formation of valid information about behavior-regulating norms. Humans, differently from other members of the animal kingdom, are "free" to act in contravention to a norm, but this will not affect the body of accumulated information. KANT (1904, p59):

"If we observe our own attitude in violating a duty, we will discover that we do not want the corresponding maxim to become a general law, *which we will feel impossible to do*, but rather that the contrary maxim should remain a general law instead; we only take the liberty of making an *exception* (only for this one time) for our advantage or to follow our inclination" (2d. emphasis in the original, 1st. emphasis added—according to the physical-science interpretation, the impossibility is due to the irreversible engraving of the belief in the rightness of the moral law which is being upheld).

Evolution of morals

These considerations lead to the work of Friedrich HAYEK (1982, 1988). Like DARWIN before him in the area of organic life, HAYEK ascertained the fact of evolution in the realm of human society. Culture,

civilization, i.e., the *order* in human society, is not due to rational planning but to "a process of selection [of rules and traditions] guided not by reason but by success", and: "Cultural selection is not a rational process; it is not guided by but it creates reason" (HAYEK 1982, pp III–166). HAYEK called this order the "extended order" of society, to differentiate it from the "order of organization", rationally created for the attainment of purposes. The main difference between the two orders is that the latter is governed by objective command-like rules intended to achieve some purposes, whereas the extended order is based on *general, purposeless, or object-unrelated* laws of behavior, which also HAYEK considers to be *moral norms*; these norms, in the same way as the general laws of nature for the behavior of non-rational beings, establish the limiting ordering frame for human action, that is, for the selection of aims, and for the actions intended to achieve them. HAYEK here confirms the main tenet of the proposed explanation, that moral norms do not, *cannot* originate in reason, in our faculty of objective cognition, as purposeful design, but are the product of a trial-and-error mechanism "guided by success".

The ascertainment of the fact of moral evolution led HAYEK to what are considered to be some of the most profound studies yet achieved of the social, economic, and political institutions of mankind. The present inquiry seeks a *reduction* of the same fact to the stratum of behavioral physiology of human individuals, the subject of the basic hypothesis proposed in Section 3. Through this hypothesis it becomes possible to extend the "carrying paradigm of evolution" (MOHR 1987, p20) to the cultural–moral level of human behavior; for, according to this hypothesis, moral beliefs, those which are upheld by the "categorical imperative", possess distinguishing qualities which, from an evolutionary viewpoint, are very similar to those possessed by biological organisms.

Once created by a mutation, biological organisms have no choice: they will uphold their constitutional characteristics with unyielding stiffness in the fight for survival, and succeed, or die out in the process; this is called "ontogenetic rigidity" in evolutionary theory, the rigidity being paramount for the accomplishment of evolution itself. The coming into being of moral beliefs cannot be said to be a product of chance, as are the mutations of biological organisms. However, as seen in Section 4, it is impossible to predict with certainty the success or insuccess of moral norms and institutions. Thus, whenever a change in the physical, technical or economic environment makes a change of norms or institutions to appear necessary, there normally arise various proposals that will have to fight for survival, exactly like biological organisms. The irreversible imprinting proposed in the basic hypothesis ensures the required rigidity for the beliefs to withstand in their fight for survival. It is thus possible to speak of an "*ontomoral rigidity*" pertaining to moral beliefs. This rigidity, and the irreversible imprinting, can be considered, as seen, as a physical-science interpretation of KANT's Categorical Imperative.

Those biological organisms which prevail in the fight for survival will induce a change in the overall bioma, the community of living organisms, which will now be better adapted to a changed physical reality. This is called "phylogenetic plasticity" in evolutionary theory. It is evident that the same is true for the cultural norms and institutions which are carried by the irreversible beliefs. Once these beliefs are shared by a sufficiently large number of people, or are carried by a group of people sufficiently strong to impose their creed, the respective norms or institutions come to be implemented in practice, and only then, through their effective implementation, will it be possible to ascertain their validity, i.e., to know if the new extended order resulting from the change is, or not, satisfactory. The beliefs in the unadapted norms and institutions will "die out", as if they were unadapted biological organisms, as exemplified by some recently overcome ideologies—a clear description of the "dying-out" process is given in Vaclav HAVEL's essay "The Power of the Powerless" (1991). There follows a build-up of "order on order" through which, exactly like the increase of information stored in the genome, an increase of information will occur about the norms and institutions which are more conducive to a satisfactory social order. It is thus possible to speak also about a "*phylomoral plasticity*" pertaining to moral beliefs.

Through the continuous build-up of order from order bestowed upon the information-holding genome, evolution has succeeded in creating a new information-holding apparatus capable of premeditated decisions for action: the human mind. This apparatus is now creating its own kind of order through a new evolutionary mechanism, the "proximate evolution" which distinguishes the human species (the "ultimate" being the genetic mutation-and-selection process, the only one capable of *creating* information). It is this proximate, or secondary evolution of moral norms, formed at the level of behavioral biology of individuals, which proceeds at

the same pace than the cultural evolution of human society, as *both are one-and-the-same phenomenon*.

The order of living beings *previous to the appearance of homo sapiens* was established directly by the general laws of nature which control, and establish limitations to, the *automatic* teleonomic actions of the living beings which compose it: a great manifold of individual purposeful (teleonomic) behaviors, ordered by general laws. Evolution of this order is the evolution of the information stored in the genome, which includes the programs for automatic instinctual behavior. With the emergence of homo sapiens, *the human behavior instinct* with its *universal behavioral program* was instrumental for the emergence of a new order, overimposed on the pre-existing order of non-rational life: the *extended order* of human society defined by Friedrich HAYEK. Qua order, it is also based on *general laws*, which, however, are not anymore immutable like the general physical laws of nature, but are now themselves subject to evolution, following the mechanism, which has been tentatively described, the *biological mechanism* which warrants the formation of *general, order-creating ethical laws*, the laws which provide the framework for the selection of values and attainment of aims of the new species.

The information-content of the order produced by moral evolution is the sum total of all those beliefs which have prevailed, that is, the set of all those "learnt rules" and "traditions" which, according to HAYEK, compose the "extended order" of society. Although their principal repository is human memory, they are also contained in documents which represent the accumulated wisdom of the ages, such as the Bible, the Koran, Roman Law, the Magna Charta, and the U.S. Constitution and Bill of Rights. They also include all the legal codes and precedents of private and public law.

The described evolutionary mechanism thus offers an objective principle of explanation for the cultural evolution of mankind, i.e., the historical development of human institutions and ways of behavior. One point, however, has to be stressed very emphatically: as in the theory of evolution of organic life, the success or insuccess of moral values can only be explained *ex post facto*. No prediction at all can be made of the future evolution of human history (cf. POPPER 1957). For, like the DARWINIAN principle of natural selection, the selection of moral norms is also a tautology: norms will either be selected, or they will not be selected, which is true in any case, not allowing any prediction as to which norms will prevail.

8. Abridged References to Works on Morals

No reference is made to works of moral philosophy, as this branch of knowledge tries to ascertain how we *ought* or *ought not* to behave, based on attempts to discover the meaning of "good" and "bad". The present inquiry, instead, seeks to ascertain the *formation* of morals (cf. *Terminology*).

The most important reference is Sigmund FREUD's theory of the "Super Ego" contained in "Civilization and its Discontents" ("Das Unbehagen in der Kultur"), written between 1929 and 1931 (FREUD 1982). At the time, behavioral physiology was still in its beginnings. The discoveries of this new science, which are associated with the life-work of Konrad LORENZ, were mostly made after FREUD's death, and the theory of the Super-Ego is thus based entirely on his psychological and psychoanalytical concepts and insights. FREUD's theory of the "Super-Ego" is essentially coincident with the basic hypothesis of Section 3. FREUD, in his deep insights into the human psyche, describes the *feelings* that govern the establishment of the ethical values which guide human behavior, or rather, the *decisions for action,* which are the source of this behavior. The behavioral explanation of the present inquiry is based on the hypothesis that the "Super Ego" *already exists* in the genetic constitution of the human species, in the form of the *universal behavior program*—the natural physiological apparatus called *conscience*. It contains dispositional learning programs into which, exactly as FREUD describes it, objectively known ethical values, together with associated feelings which permit to recognize them as "good" and "bad", are engraved through the participation of the individual in peer groups representative of the cultural environment (HARRIS 1998). And exactly as surmised by FREUD, the "internalization" of those values and feelings which characterizes the formation of the Super Ego may be considered to correspond with the "final identification" described in Section 5. Inasmuch as the "internalized" values are those originally transmitted by the parents and authorities, they become vehicles of tradition, introducing an evolutionary element in FREUD's theory, a view which is stressed by the psychoanalyst Erik ERIKSON (1997). Evolution proper, however, as described in Section 7, is missing.

Many authors have been concerned with the formation of morals. It is not possible to refer to all of them, and only some references considered representative are included in this summary. They are, in the field of biology, Edward O. WILSON, Robert TRIV-

ERS, Richard ALEXANDER, Richard DAWKINS, and also Robert RICHARDS, as representative of "evolutionary ethics". In the field of Psychology, authors include Jean PIAGET, Lawrence KOHLBERG, Erik ERIKSON, Leda COSMIDES and John TOOBY, and Judith HARRIS.

Edward O. WILSON (1975, 1978) is the main originator of the science of sociobiology, which today is considered a paradigm for the study of animal behavior. His work is mentioned in Section 1, and the present inquiry may be considered an attempt to extend the sociobiological paradigm to human behavior.

Robert TRIVERS (1985) and Richard ALEXANDER (1987) have tried to explain the evolution of morality within the established sociobiological paradigm, without trying to discover a phylogenetically developed, specifically human behavioral program that could explain the process of transmission, recognition, and internalization of objectively defined moral values. Without such a mechanism their attempt to explain the formation of morals collapses, as did WILSON's, as it reaches the apparently unsurmountable obstacle of human "free will" (cf. Section 7).

Richard DAWKINS (1989, p3): "…if you wish, as I do, to build a society in which individuals cooperate generously and unselfishly towards a common good, you can expect little help from biological nature. Let us try to *teach* generosity and altruism, because we are born selfish. Let us understand what our selfish genes are up to, because we may then have the chance to upset their designs…" Either, as Dawkins seems to presuppose, "we" are something different from our genes, which is impossible, or this sentence implies the same circularity pointed out in the discussion of WILSON (cf. "Introduction").

Robert RICHARDS' (1987, pp623–624) main justifying argument for his "evolutionary ethics" says that, because of "man's constitution as an altruist", he *ought to* behave like an altruist: "*…the evidence shows that evolution has, as a matter of fact, constructed human beings to act for the community good; but to act for the community good is what we mean by being moral. Since, therefore, human beings are moral beings—an unavoidable condition produced by evolution—each ought to act for the community good"*. (Italics in the original). Still, humans must be considered to be free to decide whether their genetic constitution should be accepted, or not, as a moral validation criterion for their behavior and acts. The "ought to" does not follow.

It is interesting to note the efforts that many scientists and philosophers have devoted to circumvent, in circuitous ways, the iron-clad naturalistic fallacy, and attempt to find some way in which the rightness or falseness of moral values could be objectively demonstrated. RICHARDS, for instance, dedicates about 16 pages of his "A Defense of Evolutionary Ethics" to an utterly ineffectual demonstration that "The Naturalistic Fallacy Describes No Fallacy". Even Edward O. WILSON, in his work "Consilience" (1998, p273) says: "…we do not have to put moral reasoning in a special category, and use transcendental premises, because the posing of the naturalistic fallacy is itself a fallacy. For if *ought* is not *is*, what is? To translate *is* into *ought* makes sense if we attend to the objective meaning of ethical precepts". This seems to be the most elegant demonstration so far, that "ought" is "is". Nevertheless, as shown in Section 4, the demonstration of the effect of a new or changed moral value on the *extended order* of human society is a practical impossibility. But it is not necessary to spend so much valuable time in trying to circumvent the naturalistic fallacy, because the nature of moral values, as shown in the present inquiry, is *evolutionary* (cf. Section 7). The evolution of the social order ascertained by HAYEK, and the biological interpretation here given, yield a most satisfactory explanation of the formation of moral values. No prediction can be made of their success, in the same way than no prediction can be made of the success of organisms in genetic evolution.

Psychologists interested in morals, like Jean PIAGET (1997) and Lawrence KOHLBERG (1980, 1981), do not try to discover the biological basis for the formation of morals, nor do they refer to the related biological texts. They attempt, instead, to find out, based mainly on observations and interviews, how morality develops in human beings, and to define stages of moral development.

Jean PIAGET, in his "The Moral Judgement of the Child", written in the 1920s, expounds the results and conclusions of a great number of interviews conducted with children aged between 6 and 12 years. His results allow him to establish the existence of three stages, or "great periods in the development of the sense of justice in the child": one period, lasting up to the age of 7–8, during which justice is subordinated to adult authority; a period approximately between 8–11 years, of "progressive equalitarianism"; and the last, setting in between 11–12, during which equalitarian justice is tempered by a sense of equity. He concludes that there exist two opposing moralities: the "ethics of authority, which is that of duty and obedience", and the "ethics of mutual respect, which is that of good (as opposed to duty), and of autonomy", and "leads to the development of equality", the source of which is "solidarity between

equals" (PIAGET 1997, p315). These findings are consistent with the biological "imprinting" hypothesis, since an inborn receptive mechanism must be present in the children's minds to allow the formation of the described moralities. It is also consistent with Judith HARRIS' theory of group socialization (cf. Section 5).

Lawrence KOHLBERG's aim is much broader than PIAGET's. He describes six "maturational" stages of moral development, defined, not as the stages of PIAGET, in terms of mind development, but as a successively more perfect conception of morality. The "highest" is the 6th stage, "*universal ethical-principle orientation*", which incorporates the Platonic vision of the "philosophical knowledge or intuition of the ideal form of the good, not correct opinion or acceptance of conventional beliefs". (KOHLBERG 1980, pp455–456). His theory is a mixture of moral philosophy and psychology, and his ideas are not always expressed with full precision and clarity, which makes their understanding difficult. This probably originates from his acknowledged rejection of the naturalistic fallacy: "…any conception of what moral judgement ought to be must rest on an adequate conception of what is". (KOHLBERG 1980, p67). Besides, his stages are mostly conceived with the aim of moral education in mind, for which it is necessary to base them on some interpretation of the meaning of "good" and "bad". The author could not find in KOHLBERG's works any reference to the biological–genetic basis for the formation of morals.

Anthropologist John TOOBY and psychologist Leda COSMIDES (1992) are prominent among the creators of evolutionary psychology. The proposition of this science related to the evolutionary creation of "modules" in the brain, corresponding to specialized reasoning devices for social exchange, has been put in doubt, as inconsistent with the findings of brain research, by brain scientists like Jaak PANKSEPP (PANKSEPP/PANKSEPP 2000), who view moral development, as in the present inquiry, as interactions between objective concepts formed in the "general computation device" of the neocortex, and motivational emotions generated in the ancient subcortical areas of the brain. As to their ideas on the evolution of instinctual behavior, cf. Section 1 and footnote 1. Their "*Wason selection task*" experiment (COSMIDES 1989), hailed by Steven PINKER as the discovery of a "cheater-detector algorithm" (PINKER 1999), may be more simply interpreted in terms of the greater or lesser capacity for formal logic reasoning of the tested college students.

Judith HARRIS' "group socialization theory", exposed in her work "The Nurture Assumption" (HARRIS 1998), which, as Steven PINKER, in his Foreword to HARRIS' work predicts, "will come to be seen as a turning point in the history of psychology", demonstrates that it is the association of children and teenagers in peer groups, which determines the moral values and norms that they will follow. The notion that the imprinting of moral values in the consciences of human individuals is a phenomenon which occurs simultaneously in the strata of groups and individuals is a central thesis of the present inquiry, and HARRIS' theory has been made an integral part of it, as exposed in Section 5.

9. The Role of Emotions: Damasio's "Somatic Marker" Hypothesis

If we say that we believe in certain values and that these beliefs govern our decisions, we silently assume that we possess the ability of generating certain feelings which will dispose us to act in ways consistent with the beliefs. What ultimately triggers our decisions to act, are emotions. This intimate knowledge about what drives our actions is probably as old as humanity, but like many other plain, obvious, self-evident facts, it has been masked to scientific insight by the exaltation of reason as the sole guide of human behavior which we owe to the "age of rationality" of the last centuries and its representative philosophers. It is yet another manifestation of the "fundamental error of category" pointed out by Gilbert RYLE, the fundamental senselessness of the dichotomic way of conceiving human nature discussed earlier.

The same basic insight motivates the work "Descartes' Error" by Antonio DAMASIO (1994). His approach to the problem, however, as a neurologist and brain scientist, originates from the stratum of neural physiology. DAMASIO departs from the observation of puzzling changes of behavior characteristic of patients suffering the consequences of a certain type of brain lesions. The area affected is the ventromedial frontal region of the brain. A much-commented case was that of the "landmark patient" Phineas GAGE who, on 13 September 1848, in New England, got his skull perforated by a rod. As a result of this accident, according to a modern reconstitution of the lesion (performed using neuroimaging techniques on his skull, which had been preserved) (Hanna DAMASIO et al. 1994), the ventromedial sectors of both cerebral hemispheres were destroyed. According to the report of John HARLOW, Phineas

GAGE's physician, in a very short time of about two months after the accident GAGE appeared to be completely recovered, except for one striking personality trait: he had taken leave of his sense of responsibility and could not be trusted anymore to keep his commitments—in other words, he had lost his capacity of making moral decisions. His employers, who had deemed him "the most efficient and capable man in their employ", now had to dismiss him.

DAMASIO goes on to compare the behavior of GAGE with that of several modern patients with similar brain damage, invariably finding the same behavioral changes.

DAMASIO thus came to the conclusion that what he calls "rational" behavior, the kind of behavior which enables us to perform adequately in our social environment, must somehow depend on a mechanism whereby the options of behavior which are open to us and which we know objectively, are brought into relation with emotions or feelings associated with the values "good" and "bad", which in turn generate the "do's" and "don'ts" characteristic of moral ("rational") behavioral decisions.

DAMASIO, following William JAMES, associates emotions at the human level directly with the "body" or "somatic" states that go with those emotions[9]. DAMASIO defines "emotions" as being manifestations of such body states, which we perceive as feelings. Feelings and emotions are not something which can be objectively formulated or quantitatively determined. Scientific measurements related to feelings must thus rely on indirect evidence, like skin electric conductance in polygraph tests.

In spite of this difficulty of grasping their elusive nature, feelings and emotions are what ultimately drives us. DAMASIO has tried to account for this fact by postulating his hypothesis of "somatic markers". These "markers" are a complex system of neural connections between the neocortical cognitive and the subcortical body regulating structures of the brain. The somatic markers (DAMASIO did not state it in this way) associate certain objectively known options for action with certain, also objectively known, reference values previously acquired by learning. The markers give rise to particular body states associated with the reference values which we associate with and categorize as "feelings" and "emotions". The "somatic marker" apparatus operates in the decision-making process.

As described in DAMASIO's book, the "somatic marker" hypothesis was tested. In one test it was conclusively demonstrated that patients with frontal lobe damage failed to produce any skin-conductance response to the display of disturbing images interspersed with neutral ones, contrasting with the control group. These patients were later able to describe in detail the emotions associated with the pictures they saw but, as inferred from the total lack of skin conductance responses, were not able to *feel* them. In another very ingenuous test based on a specially devised card game in which risk-taking was involved, it was demonstrated that the same patients lacked the ability of learning from the unfavorable results associated with the use of certain card decks. The game was so organized as to permit only the guidance of hunches and not any calculation of gains and losses. The only possible outcome predictor was thus the formation, after a certain playing time, of somatic markers. This was confirmed when the test was repeated while the subjects were hooked to a polygraph. After some time of playing normal subjects showed anticipatory responses before turning cards, whereas the patients did not. This is interpreted in the sense that they lacked the capacity to generate feelings relating the action they were about to take (the turning of a card) with a reference value "good" or "bad" created by the experience gained in the initial playing time.

It can thus be presumed that the learning process postulated as the "imprinting of beliefs" at the level of behavioral physiology consists, at the deeper neurophysiological level, of the formation of somatic markers postulated in the theory of Antonio DAMASIO. Both phenomena are one-and-the-same, but described at different strata or levels of integration of that category of being, the human species. The operation of DAMASIO's "somatic markers" can thus be equated with KANT's "imperatives", and it can be predicted with reasonable certainty that further neuropsychological investigations will reveal the existence of two basic kinds of markers:

- Flexible or changeable markers embodying the hypothetical imperatives associated with the reversible beliefs characteristic of objective knowledge;
- Inflexible, unchangeable markers embodying the categorical imperative associated with the irreversible beliefs characteristic of ethical knowledge.

By combining the theory of belief imprinting with DAMASIO's "somatic marker" hypothesis it becomes possible to formulate an operational structure of *the decision-making process*, represented in the diagram of Figure 2, the description of which follows.

The main activity of life, from which evolution arises, is problem solving. What initially presents itself is always a problem, which is recognized as

such by our genetically developed cognition apparatus, and the decision process leads to the action to solve it.

The process starts with an objective critical analysis, resulting in the statement of the problem. What follows is an initial ethical analysis to answer the question: "should I, or should I not, accept responsibility and commit myself to solve the problem?" (whose mechanism is identical to the ethical analysis described below). If yes, there follows an objective (heteronomic) analysis guided by hypothetical somatic markers (imperatives) corresponding to reversible beliefs, which leads to the statement of the available options to solve the problem.

These options are then subjected to an autonomic ethical analysis. They are brought into relation with objectively known values: on the one hand are instinctual reference values (which continue to underlie the apparatus conscience), resulting in a preview of pleasure (good) or pain (bad); on the other are the moral reference values, from which derive the previews satisfaction (good) or remorse (bad). The moral values are overimposed on the instinctual ones and are related to categorical somatic markers (imperatives). The conjunction of all these factors will generate emotional states perceived as composite feelings, whose intensity will increase until the moment of decision. As seen in Section 3, there probably exist threshold intensities which trigger the decision to accept or reject the alternative options for action, which fulfil the role of the Innate Release Mechanism (IRM) of our non-rational relatives.

Once a decision has been reached, there still may remain to be done an objective optimization analysis, guided by heteronomic somatic markers (imperatives), leading to the final statement of the problem solution, and finally to action.

It may be noted that the complete process as depicted, will probably occur only for serious, momentous decisions. Ordinary decisions, which very often occur within an already accepted ethical reference frame, may occur "on the spur of the moment", being guided exclusively by heteronomic somatic markers.

DAMASIO's main conclusion is that what he calls "rational" behavior is guided by feelings. He thus rejects as untrue the contention that "pure reason" would be the sole guide of human behavior, and this is why he called his book "Descartes' Error". The exaltation of reason as guide of behavior is mainly due to DESCARTES, but other "pure reasoners" like KANT are also included in DAMASIO's critique. From the present inquiry, however, it will be recognized that KANT's work cannot be rejected out of hand but that, to the contrary, it represents a fundamental step in the advance of knowledge about ourselves.

It must be noted that DAMASIO's "somatic marker" hypothesis is put in doubt by some neurobiologists, who argue that the empirical support is not clear, and that a subcortical theory would be more adequate (Jaak PANKSEPP, personal communication; cf. also PANKSEPP/PANKSEPP 2000). Nevertheless, it cannot be put in doubt that human beings are biological units, from the group level down, and that there must be an exact and simultaneous topological correspondence of events affecting the unity at the group, behavioral, and neurophysiological levels. Thus, if the cited empirical tests of DAMASIO's hypothesis lack sufficient conviction, a different, or a variation, of DAMASIO's theory will explain those events at the neurophysiological level, and will be as consistent with the imprinting-of-beliefs hypothesis as DAMASIO's "somatic markers" hypothesis is.

10. Summary and Conclusions

The described inquiry departs from the fact that the human species is a new category of existence which, although being included in nature and governed by natural laws, possesses a new upper stratum of cognitive faculties which distinguishes it from all the other species of the animal kingdom. These faculties reside in the cognitive apparati of *objective* and *moral cognition*, which control human behavior. Since human attitudes and behavior are *non-automatic*, the natural laws by which they are governed must explain, as stated in the Introduction, how the *information* concerning behavior-governing rules and norms originates. This is done, as postulated in the attempted explanation, through the phylogenetically evolved *human behavior instinct*, which constitutes, in effect, a *universal behavioral program* possessed only by humans.

The *human behavior instinct* operates, as proposed in the basic hypothesis, through the process of "*imprinting*", by affixing action-governing feelings to beliefs in the truth or rightfulness of objectively known norms and values. These feelings govern *the human decision process,* which in humans acts in lieu of, and thus controls to a large extent, the release mechanism that activates the genetically preprogrammed actions of the automatic instincts. The hypothesis also states that the imprinting of beliefs in *heteronomic* rules is *reversible*, whereas the imprinting of beliefs in *autonomic* rules is *irreversible*. Changes of heteronomic rules in the realm of objective cogni-

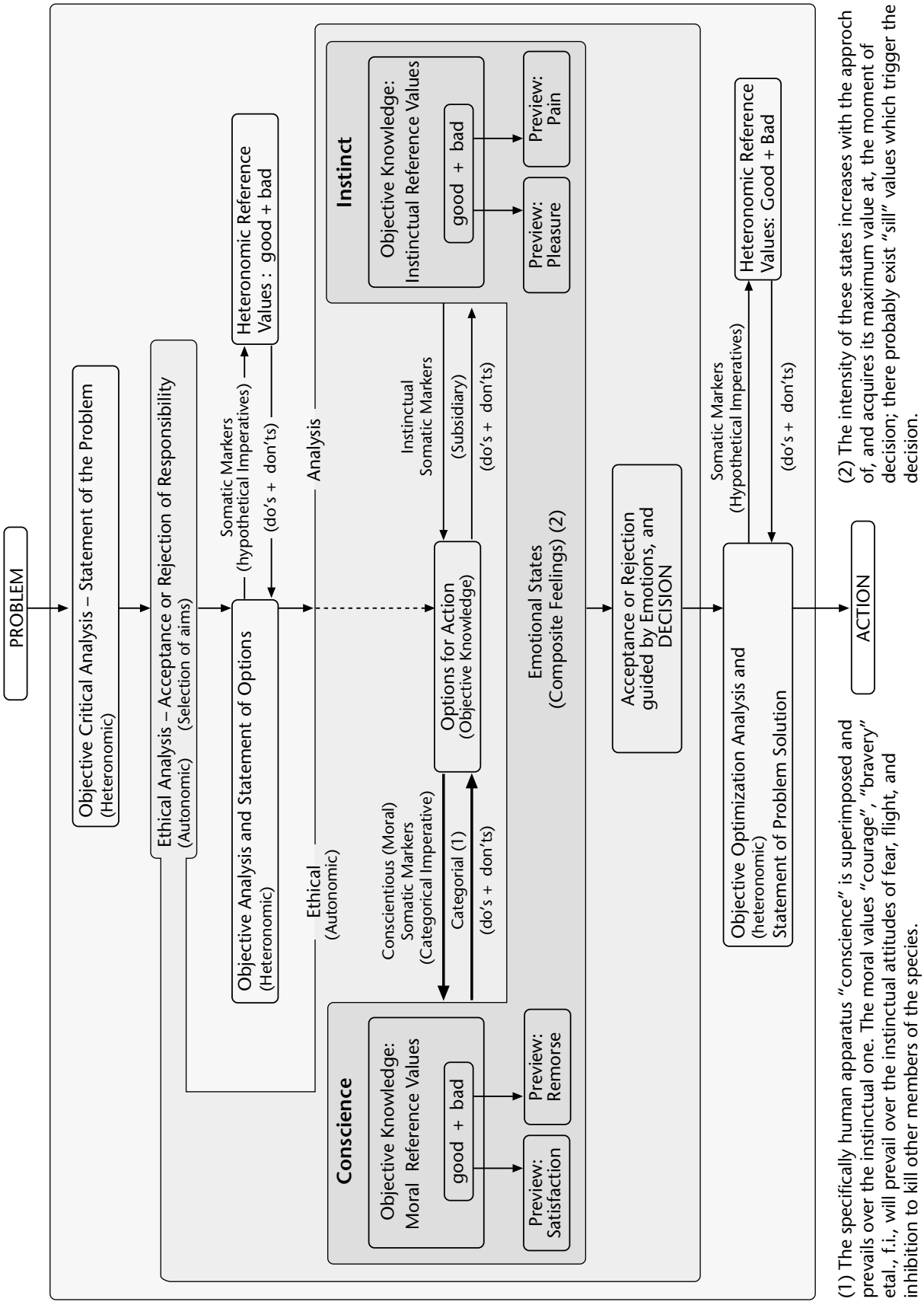

Figure 2: The Decision Process

tion occur if and when they are falsified by new or better rules, and the respective belief is reversed. Autonomic rules, instead, govern the field of moral cognition, where predictions are impossible and where the validity of a rule can only be demonstrated after its implementation, ex-post-facto. Phylogenetic adaptation thus caused the irreversibility of the imprinting of beliefs in this domain, and the *ontomoral rigidity* displayed by their carriers. This permits to conceive and describe the process of *Evolution of Morals*, which determines the changes of autonomic rules and values. Beliefs in moral values carried with ontomoral stiffness by large groups of people will come to be implemented and tried out in practice, submitted to the merciless trial of their adaptation to reality as if they were biological organisms, and will survive or die out in this process, which governs human history.

Author's address

Carlos Stegmann, Rua dos Pessegueiros 101, 05673-010 Sao Paulo, S.P., Brazil.
Email: stegmann@amcham.com.br

It thus became possible to extend the findings of behavioral science to the human species, in a way which avoids the logical incongruities which arise from the fact that moral decisions by individuals and groups are left out from the explanation of human behavior so far given. It also avoids the insolvable problems confronted by evolutionary ethics when trying to construct a system of morals based on our innate instinctual propensities built up by genetic evolution, problems which have been pointed out in Paul Lawrence FARBER's "The Temptations of Evolutionary Ethics" (1994), and by HESCHL (1998). Authors like Irinaeus EIBL-EIBESFELDT (1984) and Robert J. RICHARDS (1987) say that "we should" follow those propensities—but it is the uttering of "we should" itself, i.e., *the moral decisions themselves*, which must be explained by an objective theory of morals. This was the object of the present inquiry.

Notes

1 HESCHL totally disagrees with the statement of TOOBY and COSMIDES, that "The fact that our behavioral mechanism operates successfully under the altered modern conditions is a purely secondary consequence of its Pleistocene-forged design", which he considers nonsensical. Whenever the *human behavior instinct*, as formulated in the present theory, was created by mutation-and-selection, it would be capable to adapt human behavior not only to conditions prevailing in the Pleistocene, but to *any* kind of conditions.

2 "Moral attitudes" is an expression which is consistent with the present theory. Other authors (e.g., PANKSEPP/PANKSEPP 2000) refer to attitudes like those of ROSE and LEWONTIN as "political biases".

3 The determination of the meaning of words is subject to what Hans ALBERT calls "the Münchhausen trilemma": infinite regression, circularity, or interruption of proceedings (ALBERT 1984). This implies, as Karl POPPER (1976) argues, that the "essential" meaning of words can never be conclusive, but also that this "essential" meaning of words is epistemologically unimportant, whereas the meaning of theories is epistemologically all-important.

4 In his 1997 paper OESER proposes a second-stage evolution attuned to the cognitive faculties of the human species, in coincidence with one of the basic tenets of the present inquiry.

5 The analytic faculty resides in phylogenetically evolved rules of logic that operate with *abstract* concepts or categories. These are, originally, representations of things perceived in the environment, but the analytic apparatus converts them into the generic form of *abstractions*. As such, they cease to have a *direct* relation to any individual or concrete object of reality, but can now be operated with the rules of logic, yielding deductive results which are always certain. We humans have an irresistible compulsion, an emotion generated in the subcortical areas of the brain, to apply our analytic faculty to the comprehension of reality. We will apply the abstract categories and the rules of logic to *real* objects, deriving results which are called *synthetic* judgements, or statements about what happens in the real world, which will always be uncertain; but in spite of this uncertainty, these judgements usually represent sensible approximations, and caused the enormous survival advantage and the rapid spread of homo sapiens. (This is based on CARNAP 1995). The analytic faculty of the human mind/brain has been created by the mutation-and-selection mechanism of evolution, the genome being its sole repository (HESCHL 1998). As created in the genome, this faculty can be ontogenetically expanded. Thus, the rules of logic can be, and have been, vastly expanded into abstract sciences like mathematics and geometry. The difference between the *analytic* deductions, which are always certain, and the *synthetic* judgements, which are always uncertain, however, will be always maintained. Albert EINSTEIN expressed this very poignantly in his essay "Geometry and Experience" (1982): "As far as the propositions of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality".

6 "For the task of a great philosopher is not to reach correct conclusions, but to be the voice of his time, to put the Zeitgeist ("das Weltgefühl") of his epoch into a system of thought". (Friedell 1976, p500, translated from German). From the objective point of view of the imprinting mechanism it is indifferent if the philosopher, social reformer, religious or political leader is René DESCARTES (to whom FRIEDELL's comment refers) or Osama BIN LADEN, Winston CHURCHILL, or Adolf HITLER. These persons and their follow-

ers represent the "conspecifics" to which the beliefs in the rightness of certain moral convictions are attached (cf. Section 3).

7 The *general* character of moral laws preconditions the possibility of statistics in the social field. It permits the formation of spontaneous statistical values such as prices, and also the mathematical formulation of trends, which constitute a fundamental input for the activity of planning (STEGMANN 1994) .

8 The "ratiomorph" faculties, and the conventional instincts, which humans also possess, are directly comparable to those of irrational living organisms.

9 "Our natural way of thinking about … emotions is that mental perception of some fact excites the mental affection called the emotion, and that this latter state of mind gives rise to the bodily expression. My theory, on the contrary, is that *the bodily changes follow directly the perception of the exciting fact, and that our feeling of the same changes as they occur* IS *the emotion".* (JAMES 1952, p743, emphases in the original).

References

Albert, H. (1984) Kritische Vernunft und Menschliche Praxis. Philipp Reclam: Stuttgart.

Alexander, R. (1987) The Biology of Moral Systems. De Gruyter: Hawthorne NY.

Carnap, R. (1995) An introduction to the philosophy of science. Edited by Martin Gardner. Dover Publications: Mineola NY. Originally published in 1966.

Cosmides, L. (1989) The logic of social exchange: Has natural selection shaped human reason? Studies with Wason Selection Task. Cognition 31:187–276.

Cosmides, L./Tooby, J. (1992) The adapted mind: Evolutionary psychology and the generation of culture. Oxford University Press: New York.

Damasio, A. (1994) Descartes' error. G. P. Putnam's Sons: New York.

Damasio, H. et al. (1994) The return of Phineas Gage: Clues about the brain from the skull of a famous patient. Science 264:1102–1105.

Dawkins, R. (1989) The selfish gene. Oxford University Press: New York.

Eibl-Eibesfeldt, I. (1984) Krieg und Frieden aus der Sicht der Verhaltensforschung. Piper: Munich.

Einstein, A. (1982) Geometry and experience. In: Ideas and Opinions. Crown Publishers: New York, pp. 232–245. Originally published in 1954.

Erikson, E. H. (1997) The life cycle completed. W. W. Norton: New York.

Freud, S. (1982) Studienausgabe Band IX. Fischer Taschenbuch Verlag: Frankfurt Main. First editions: Die 'kulturelle' Sexualmoral und die Moderne Nervosität (1908); Zeitgemäßes über Krieg und Tod (1915); Das Unbehagen in der Kultur (1930).

Friedell, E. (1976) Kulturgeschichte der Neuzeit. Deutscher Taschenbuch Verlag: Munich. Originally published in 1927.

Hamilton, W. D. (1964) The genetical evolution of social behaviour. Journal of Theoretical Biology 7:1–16;17–52.

Harris, J. R. (1998) The nurture assumption. The Free Press: New York.

Havel, V. (1991) The power of the powerless. In: Open Letters. Selected Prose 1965–1990. Faber and Faber: London, Boston, pp. 125–214. Originally published in 1978.

Hayek, F. A. v. (1982) Law, legislation and liberty. Complete edition. Routledge & Kegan Paul: London. First editions: Vol. 1, Rules and Order (1973) Routledge & Kegan Paul: London; Vol. 2, The Mirage of Social Justice (1976) University of Chicago Press: Chicago. Vol. 3, The Political Order of a Free People (1979) Institute of Economic Affairs: London.

Hayek, F. A. v. (1985) Die Überheblichkeit der Vernunft: Standortsbestimmung des Generalthemas Evolution und die Zukunft der Menschheit. Eröffnungsvortrag zum Europäischen Forum Alpbach 1985. Österreichisches College: Wien.

Hayek, F. A. v. (1988) The fatal conceit: The errors of socialism. Routledge: London.

Heschl, A. (1998) Das Intelligente Genom: Über die Entstehung des menschlichen Geistes durch Mutation und Selektion. Springer Verlag: Berlin, Heidelberg.

James, W. (1952) The principles of psychology. Encyclopædia Britannica: Chicago. Originally published in 1890.

Kant, I. (1904) Grundlegung zur Metaphysik der Sitten. Philipp Reclam Jr: Leipzig. Second edition revised by Kant originally published in 1786.

Kant, I. (1966) Kritik der reinen Vernunft (Second Edition). Philipp Reclam Jr: Leipzig. German original published in 1776.

Koestler, A. (1967) The ghost in the machine. Arcana Books: London.

Kohlberg, L. (1980) Stages of moral development as a basis for moral education. In: Brenda M. (ed) Moral development, moral education, and Kohlberg. Religious Education Press: Birmingham AL, pp. 15–100 and 455–471.

Kohlberg, L. (1981) The meaning and measurement of moral development. Clark University Press: Worcester.

Lorenz, K. (1963) Das sogenannte Böse: Zur Naturgeschichte der Aggression. Deutscher Taschenbuch Verlag: Munich.

Lorenz, K. (1973) Die Rückseite des Spiegels. Deutscher Taschenbuch Verlag: Munich.

Lorenz, K. (1978) Vergleichende Verhaltensforschung. Springer Verlag: Wien, New York.

Lorenz, K. (1983) Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie. In: Lorenz, K./Wuketits, F. M. (eds) Die Evolution des denkens. Piper: München, pp. 95–125. Originally published in 1941.

Lorenz, K. (1992) Die Naturwissenschaft vom Menschen: Eine Einführung in die vergleichende Verhaltensforschung. Das "russische Manuskript". Piper: München.

Mohr, H. (1987) Natur und Moral. Ethik in der Biologie. Wissenschaftliche Buchgesellschaft Darmstadt: Darmstadt.

Oeser, E. (1983) Die Evolution der wissenschaftlichen Methode. In: Lorenz, K./Wuketits. F. M. (eds) Die Evolution des Denkens. Piper: München, pp. 263–300.

Oeser, E. (1996) Evolutionary epistemology as a self-referential research program of natural science. Evolution and Cognition 2(1):16–21.

Oeser, E. (1997) The two-stage model of evolutionary epistemology. Evolution and Cognition 3(1):85–88.

Panksepp, J./Panksepp, J. B. (2000) The seven sins of evolutionary psychology. Evolution and Cognition 6(2):108–

131.
Piaget, J. (1997) The moral judgement of the child. English translation by M. Gabain. Simon & Schuster: New York. French original published in 1931.
Pinker, S. (1995) The language instinct. Harper Collins: New York. Originally published in 1994.
Pinker, S. (1999) How the mind works. Norton: New York.
Popper, K. R. (1957) The poverty of historicism. Routledge & Kegan Paul: London.
Popper, K. R. (1976) Unended quest: An intellectual autobiography. Fontana/Collins: London.
Popper, K. R. (1990) The logic of scientific discovery. Unwin Hyman: London. Originally published 1959.
Richards, R. J. (1987) Darwin and the emergence of evolutionary theories of mind and behavior. The University of Chicago Press: Chicago.
Riedl, R. (1975) Die Ordnung des Lebendigen. Piper Verlag: Munich.
Riedl, R. (1979) Biologie der Erkenntnis: Die stammesgeschichtlichen Grundlagen der Vernunft. Deutscher Taschenbuch Verlag: Munich.
Ryle, G. (1949) The concept of mind. Penguin Books: London.
Rose, S./Lewontin, R.C./Kamin, L. J. (1984) Not in our genes: Biology, ideology and human nature. Penguin Books: London.
Schrödinger, E. (1944) What is life? Cambridge University Press: Cambridge.
Stegmann, C. (1994) Limits of reasonable planning. Journal of Professional Issues, A.S.C.E. 120(3):246–253.
Trivers, R. (1985) Social evolution. Benjamin & Cummings: Menlo Par CA.
Vollmer, G. (1990) Evolutionäre Erkenntnistheorie. S. Hirzel Wissenschaftliche Verlagsgesellschaft: Stuttgart.
Vollmer, G. (1996) Wollen-Können-Dürfen. In: Daecke, S. M./Bresch, C. (eds) Gut und Böse in der Evolution. Edition Universitas: Stuttgart, pp. 69–83.
Wilson, E. O. (1975) Sociobiology: The new synthesis. Harvard University Press: Cambridge MA.
Wilson, E. O. (1978) On human nature. Penguin Books: London.
Wilson, E. O. (1998) Consilience. Alfred A Knopf: New York.

Conrad Montell

On Evolution of God-Seeking Mind

An Inquiry into Why Natural Selection Would Favor Imagination and Distortion of Sensory Experience

Abstract

The earliest known products of human imagination appear to express a primordial concern and struggle with thoughts of dying and of death and mortality. I argue that the structures and processes of imagination evolved in that struggle, in response to debilitating anxieties and fearful states that would accompany an incipient awareness of mortality. Imagination evolved to find that which would make the nascent apprehension of death more bearable, to engage in a search for alternative perceptions of death: a search that was beyond the capability of the external senses. I argue that imagination evolved as flight and fight adaptations in response to debilitating fears that paralleled an emerging foreknowledge of death. Imagination, and symbolic language to express its perceptions, would eventually lead to religious behavior and the development of cultural supports. Although highly speculative, my argument draws on recent brain studies, and on anthropology, psychology, and linguistics.

Key words

Evolution, imagination, mortality, self-awareness, fear, religious behavior, language.

Setting the Stage for Imagination and Religious Behavior

"It was the experience of mystery—even if mixed with fear—that engendered religion" (EINSTEIN 1954, p11).

For early *Homo sapiens*, big-brained and naturally curious, the emergence of self-awareness and a nascent awareness of mortality (perhaps as spandrels: as unavoidable consequences of increased brain size and intelligence) surely would lead to that experience of mystery. It also would lead to a new kind of survival problem. In contrast with specific responses to specific threats, what could be an appropriate response to awareness of a pervasive threat, an unavoidable danger that was not salient in the natural environment? How could such awareness benefit survival? Feeling the presence of such a predator, where there is no possible flight or fight, might more likely incapacitate or frighten one to death. Such awareness could hardly be reproductively beneficial unless it led to some adaptation that reduced the perceived danger. But what? Swifter legs? Keener sight? Sharper teeth? Stronger arms? None of these would do. What then? Since that "predator" lurks somewhere in the brain, so too, the adaptation—as some mental structure to counter or at least mitigate that awareness.

The unique and yet unexplained aspects of human evolution are common knowledge. Among the multitude of adaptations that evolved in species, there appears to be this one set for which there is no antecedence in other species: the adaptations that form the human mind (LORENZ 1977). There appears to be a discontinuity in evolution when it comes to the human mind (DONALD 1991). "Biologically, we are just another ape. Mentally, we are a new phylum of organisms" (DEACON 1997, p23). In considering the distinct form of life that is the human mind, some might consider it to be a new kingdom (LORENZ 1977). That such adaptations evolved and flourished only in *Homo sapiens* suggests the existence of a unique stimulus in the formation period of our species. This paper focuses on that stimulus, and on evolutionary and behavioral responses to it.

We have evolved with an awareness of the world that goes beyond externally sensed reality, with an inner "sense" that creates its own reality. We have evolved with unique ways of perceiving the world, and with unique ways of passing on information to future generations, who benefit from the survival value of our behavior as well as the information in our genes (DEACON 1997; DENNETT 1995, 1978;

DONALD 1991; LORENZ 1977; MITHEN 1996; PINKER 1997). Consider certain quantitative and qualitative differences in the animal world. The difference between the brain of a fruit fly and that of a chimpanzee can be viewed as quantitative: the chimp has much more and much better of the same kind of brain material. In contrast, the difference between the brain of a chimpanzee and that of a human must be viewed as qualitative: beyond the measure of DNA, as seen in differences in behavior and cognition, we have some qualitatively different material, which other primates do not have (BRONOWSKI 1977, DONALD 1991, LORENZ 1977, MITHEN 1996). "It is as if all life evolved to a certain point, and then, in ourselves turned at a right angle and simply exploded in a different direction" (JAYNES 1976, p9). The extraordinary gap in mental performance between humans and the rest of the animal world has defied efforts to bridge it with plausible explanation. LORENZ refers to a great gulf produced by "a creative flash", a "fundamental revolution of all life brought about by the coming into existence of the human mind" (1977, p167), "utterly impenetrable to the human understanding" (p169). Part of that "fundamental revolution" can be understood, I suggest, by looking at a new kind of self-preservation behavior stemming from a threat to life that only humans have perceived.

All animals behave to survive and reproduce, and all require sensory equipment in order to gain accurate information from the environment. Indeed, in this kind of behavior, we are just like other primates. However, there is also "and not by bread alone behavior" to account for: unique behavior and a unique problem. Humans have had an awareness of a non-specific threat to life, and humans have evolved with equipment and behavior in order to cope with that perceived threat. At some evolutionary stage, proto-humans began to be aware of *self* and *other*, of time past, and of approaching time beyond the given moment. As a consequence, they eventually became aware of their mortality (a still evolving awareness), and suffered the throes of that awareness as well as that of death itself (BECKER 1973; BROWN 1959; LANGS 1996; LANGER 1982, 1972, 1967; PYSZCZYNSKI/GREENBERG/SOLOMON 1997). Other animals, whose awareness is imprisoned in present time (BRONOWSKI 1977), merely suffer the throes of death. Animals have developed brains; humans have developed additional equipment and the ability to communicate symbolically (DEACON 1997; DONALD 1991). I argue that some part of that equipment and that linguistic behavior developed in response to the stimulus of potentially debilitating fear brought on by an evolving awareness of mortality. Rather than an adaptation (what reproductive benefit is there in this awareness?), it may have been an inevitable consequence of a certain level of brain complexity. Once in place it would lead to new human behavior. "The function of the brain is to produce behavior. The function of behavior is to promote the DARWINIAN fitness of the behaver" (STADDON/ZANUTTO 1998, p242). What is true for the animal brain should also be true for that additional equipment known as the human mind.

DENNETT, shifting the mind–brain problem by referring to "animal minds", discusses the "huge difference between our minds and the minds of other species… We are also the only species with language" (DENNETT 1995, p371). Why only us? He answers by posing another question: "What varieties of thought require language?" (p371). One such variety of thought, I suggest, is religious in nature. DENNETT proposes a design structure for the ascendance of human mind which he calls "the Tower of Generate-and-Test" (p373). Here again, why would only human minds climb to the top of such a structure? DENNETT suggests the advent of tool use, but does not address the question of why only humans so used tools. He speaks of a device for lifting the brain to human heights: "the crane to end all cranes: an explorer that *does* have foresight, that can see beyond the immediate neighborhood of options" (p379). Again, the question as to what need led only human brains to look "beyond the immediate neighborhood of options?" is unanswered. DENNETT disagrees with those who refer to human "mysteries" such as free will: human puzzlement that cannot be solved. My thesis suggests at least a partial answer to the question of human uniqueness and a solution to one of the mysteries: the development of religious behavior. Whether we call it *mind* or *brain*, the human intellect-imagination system evolved to engage in behavior that cannot simply be described in terms of physical survival. Part of this "and not by bread alone" behavior is religious in nature. "As every creature and even every living tissue responds to stress with heightened activity, so the mind meets the challenge its own evolution has created by a radical deepening of religious feeling and dawning of religious ideas" (LANGER 1982, p110).

My thesis does not address the complexity of needs served by organized religion, the moral aspects of religious activity, or other aspects of the ubiquitous *mind*. The focus is not on whether religious behavior is adaptive in the modern world.

Rather, the focus is on imagination as a possible adaptive response for early *Homo sapiens* to that cardinal human fear: mortality (LEYHAUSEN 1973), and on the associated memory devices essential for storing the products of imagination (LANGER 1982, 1972). Apprehension of death developed as a free fear: the sensing of a danger that cannot be avoided or fled from (LEYHAUSEN 1973; LANGS 1996). Having such apprehension, "we die a thousand deaths, that is the price we pay for living a thousand lives" (BRONOWSKI 1977, p25). When and to what extent this apprehension became *conscious* (in the ordinary murky sense of the word) are questions beyond the scope of this paper. This apprehension might have developed as a consequence of that prereflective consciousness SARTRE and others consider as awareness of an object and awareness that *it* is not that object (MALHOTRA 1997). I avoid modern issues of authenticity of self and self awareness: issues of whether and to what extent such self and awareness exist and are known by the individual, apart from social content (WEIGERT 1988). It seems that at least some amount of self-awareness is required to enter the state of being a self (MARTIN 1985, p3). I intend *awareness*: of self and other, and of mortality, to mean some "knowing" of these things that leads to behavior, whether or not the knowing can be squeezed into thought and expressed. Thus, this sense of *awareness* encompasses various forms of *knowing*, some of which were (and still are) ineffable: anxiety, feelings of foreboding, dread, and individual moods that find expression in some form of human behavior, including inaction (out of fear) in a situation calling for action. There is FREUD's controversial conception of a *death instinct* to consider, as well as other instinctual knowing that exists at the borderline of animal and human awareness (BROWN 1959). Considering these levels of the knowing of fear, I focus on that cardinal fear and on the potential loss of vitality that I suggest paralleled its development: "a number of factors, psychological as well as physiological in nature, at work in causing actual, concrete fears; the cardinal source (not the experienced but the essential one) of the phenomenon of fear as a whole, however, is man's mortality" (LEYHAUSEN 1973, p248).

I argue that human imagination evolved as a way of coping with that cardinal fear and its potentially debilitating consequences. This fear could not be alleviated by further evolution of the external senses. An inner sense offered an escape from a "predator" that did not appear within the physical environment. This escape mechanism quite likely developed as a distortion of sensory experience (LORENZ 1977; LEYHAUSEN 1973). This *disorder* had survival value for our species. LORENZ offers a clue in understanding such a development: "Far from hindering the investigation of the organism affected by it, a pathological disorder very often gives us the key to the understanding of how the organism works" (1977, p5). This "key", I suggest, is useful in understanding the evolution of imagination as an adaptation. Human self-awareness leading to an awareness of mortality can be considered a disorder, and just that kind of disorder that "gives us the key to understanding" how that anomalous part of the human animal came into being. "We have developed 'organs' only for those aspects of reality of which, in the interest of survival, it was imperative for our species to take account, so that selection pressure produced this particular cognitive apparatus" (LORENZ 1977, p7). Since human imagination appears to be unique, it seems reasonable to inquire as to what unique survival problems might have developed for proto-humans. I suggest that an evolving awareness of self and of death of self led to a new kind of survival problem that, in turn, led to the evolution of a new kind of "solution". As it evolved, imagination would lead to the development of "belief", a pro-attitude superimposed on information and experience, and to a new kind of behavior: religious behavior. "Its original function may have been to keep men's minds in balance with the rest of nature, but what has led to its own elaboration is a purpose it soon acquired: the denial or masking of death" (LANGER 1982, p137).

"Religious behavior" is used here in a broad sense, to include the nascent mental activity hominids, newly aware of self and mortality, might have engaged in individually and, as emerging language made possible, in small groups. With regard to investigating the sources of religious behavior, there is, of course, a great rift in human views of "mind", "soul", and individual afterlife: a largely unspoken-of dichotomy between scientists and secular academicians on the one hand, and the rest of the world on the other, between the staunch materialists (monists) and the mass of people (dualists) who feel that mind and soul exist as non-material stuff. "I suppose most people in our civilization accept some kind of dualism. They think they have both a mind and a body. But that is emphatically not the current view among the professionals in philosophy, artificial intelligence, neurobiology, and cognitive science" (SEARLE 1997, p43). Few professionals address this rift. There is some risk in doing so, especially when the different views are taken beyond academia. Our

beliefs have detrimental consequences for ourselves and others (LANGS 1996). Ancient and still active concepts such as *faith, sacred, worship*, as well as the central concept of *God*, all with little or no direct relationship to physical survival in the externally sensed world have, nonetheless, led to life and death conflict. Wars have been fought, and are even now being fought, masses of people killed, because of differences in religious belief. Undoubtedly, all this has contributed to the dearth of scientific inquiry into evolutionary sources of belief and the potential role of imagination in the development of religious activity. Many religious differences, at the rarely-exposed marrow of belief, center on what continuance there might be for an individual mind after death, and on what behavior might influence such continuance. The question appears to be as old as the mental equipment required for asking it.

Imagination: Structure and Process

Unlike the information-seeking external senses, *imagination* creates its own information: new and sometimes distorted images of the natural world. It is a basic human characteristic, more basic than intelligence, which is abundant in the animal world (BRONOWSKI 1977). As here described, *imagination* is an aspect of mind that we know by its lexical meaning: "the act or power of forming mental images of what is not actually present; the act or power of creating mental images of what has never been actually experienced, or of creating new images or ideas by combining previous experiences; creative power" (WEBSTER 1996). It is the "employment of past perceptual experience, revived as images in a present experience at the ideational level" (DREVER 1964, p130), "the process of creating objects or events without the benefit of sensory data" (CHAPLIN 1985, p221). STEPHEN speaks of the existence of *autonomous imagining*, "imagery so compelling, so powerful it can even override all demands of external reality" (1989, p56), imagery "experienced as an external, independent reality", and propose that religious experience "is grounded in the psychological reality of a special imaginative process operating outside ordinary awareness" (p212). JAMES describes that experience, "the convincingness of what it [imagination] brings to birth. Unpicturable beings are realized, and realized with an intensity almost like that of an hallucination. They determine our vital attitude as decisively as the vital attitude of lovers is determined by the habitual sense, by which each is haunted, of the other being in the world" (1936, p71).

It is useful here to distinguish between two fundamental mental attributes: intellectual and imaginative. Compared to the intellect, imagination is a more subtle mental phenomenon, seemingly impossible to quantify (ECCLES 1989). "The imaginative process is the human capacity to evoke an image or an idea in the absence of a direct perceptual stimulus" (RANGELL 1988, p63), "to make images and move them about inside one's head in new arrangements" (BRONOWSKI 1977, p24). BERES defines imagination broadly, "as the capacity to form a mental representation of an absent object, an affect, a body function, or an instinctual drive… a *process* whose *products* are images, symbols, fantasies, dreams, ideas, thoughts, and concepts" (1960, p327). DENNETT speaks of images as existing within a *phenomenal space* that can contain a god or heaven as well as a tangible object: "Phenomenal space is Mental Image Heaven, but if mental images turn out to be *real* they can reside quite comfortably in the physical space in our brains, and if they turn out not to be real, they can reside, with Santa Claus, in the logical space of fiction" (1978, p186). I suggest that for early *Homo sapiens* with emerging imagination (as for a large number of modern humans), real objects and "Santa Claus" reside together quite harmoniously.

To all this I would add that, in relation to the brain's processing of external information, imagination functions as sensory-distorting perception. To the extent that this perception leads to something new that can be shared, we might call it "creative imagination". Here, individual processes are extended to those of a social nature: to the sharing of illusions and the formation of new images as a social process. Products of imagination are qualitatively different from mere illusions, from that perversion of sense-data which might occasionally have taken place in pre-imaginative hominid brains (and in those of other animals). With the advent of imagination, illusions would increase and assume new forms and new functions. One positive function would be to divert the individual from fearful thoughts involving "self" and change. KOESTLER speaks of this function as: "the transfer of attention from the 'Now and Here' to the 'Then and There'—that is, to a plane remote from self-interest" (1964, p303). In an imaginative state, a state identified as Absolute Unitary Being, a state described in the mystical literature of the world's most ancient religions, individuals lose their sense "of discrete being, and even the difference between self and other is obliterated" (D'AQUILI/NEWBERG 1998, p195). Religious literature describes imaginative states in which indi-

viduals lose their awareness of self and with it lose their mortal fears. In such states, rather than adding to awareness, imagination acts as a filter, a curtain, or even as a screen, distorting, dimming, or obliterating awesome perceptions. In such states, imagination serves to transport or sever the individual from the sources of mortal fear. Historically, in literary theory, "it was opposed to reason and regarded as the means for attaining poetical and religious conceptions" (HOLMAN/HARMON 1992, p241).

As it first evolved, imagination, undoubtedly, would merge with older forms of illusion—in and out of dreams. "In primitive stages of hominid specialization dream may not have occurred exclusively or even mainly in sleep. For eons of human (or protohuman) existence imagination probably was entirely involuntary, as dreaming generally is today, only somewhat controllable by active or passive behavior" (LANGER 1972, p283). LANGER gives support to the view of the pioneering French psychologist Jean PHILIPPE who described imagination as a kind of biological entity: "In the complexity of our mental organization it is a sort of living cell, which maintains its life through manifold and diverse transformations" (PHILIPPE 1903, p4). Whatever it physically consists of, imagination most likely evolved with *Homo sapiens*. Expressions of it are difficult or impossible to detect from the monotonous tools and other archeological finds from the long record of *Homo erectus*, although it seems likely that at least gestation of human self-awareness had begun by the end of this period of some hundred thousand generations of big-brained and potentially aware creatures who became extinct 200,000 years ago. MITHEN, pondering how little *Homo erectus* seemed to create with his large brain, speaks of a "shuffling of the same essential ingredients" in their technology for more than a million years, with only "minor, directionless change" (1996, p123).

Imagination in Its Early Forms

The earliest artifacts that have been found to express imagination, those from the late Middle and early Upper Paleolithic periods, express religious activity having to do with death and mortality. The earliest traces of beliefs and practices are of such religious form: Neanderthal burials seventy thousand years ago and perhaps even older burials in China; elaborate Paleolithic cave art drawn in dark, tortuous, difficult to access recesses; evidence of animal worship and of rituals associated with hunted animals; and other prehistoric evidence of the struggle to understand and come to terms with individual death and glimmers of mortality (DONALD 1991; HOLMES 1996; PARRINDER 1984). In historic times we see this struggle for understanding expressed in the earliest literature, in all known cultures. These cultural products express the religious thought that seems to be the primal focus of human imagination, as we first encounter such imagination in salient human behavior (BROWN 1959; DENNETT 1995; FREUD 1950; HOCART 1954; JAMES 1936; JAYNES 1976; LANGS 1996, LANGER 1982; MITHEN 1996). I argue that the imaginative parts of *mind* were naturally selected in response to debilitations that paralleled awareness of mortality. Imagination and companion devices to process and store its products in memory evolved to mitigate that awareness, to discover offsetting information beyond the apparent horizon, to sense a more favorable reality, and thus, to make the emerging awareness of death more bearable, and to make the aware individual more fit. Although much of the prehistoric process may never be known, evidence for this function of imagination permeates history and contemporary human life. DONALD describes the universal importance of religious belief within hunter–gatherer societies, all of whom appear to have an elaborate mythological system similar in principle:

"Myth permeates and regulates daily life, channels perceptions, determines the significance of every object and event in life. Clothing, food, shelter, family—all receive their 'meaning' from myth. As a result, myths are taken with deadly seriousness: a person who violates a tribal taboo may die of fear or stress within days, or be ostracized, or put to death" (DONALD 1991, p215).

There is neuropsychological data to suggest that "human beings have no choice but to construct myths consisting of personalized power sources to explain their world" (D'AQUILI/NEWBERG 1998, p191). Supporting this, a range of cultural products reveals the primacy of mortal fears and religious hopes in diverse societies throughout time and throughout the world. Every known social group has had a religion that includes some sense of immortality or some attempt to deny the reality of death (BROWN 1959). As one well-documented example, Egypt, four thousand years ago, a society of some seven million people, devoted the bulk of its surplus and some of its essentials to the building of monuments for its Pharaohs. To prepare dead bodies for entry into an imagined next world, living bodies suffered hunger in this world. There is evidence, in the caves that housed them, that many of the hundreds

of thousands of pyramid builders and artisans labored willingly for their Pharaoh's afterlife and for their own. Today, with five billion of the world's six billion as adherents, ancient religions are alive and flourishing, 143 years after *On the Origin of Species* and their predicted demise. In a nationwide poll by The New York Times and CBS News of over a thousand teenagers, "ninety-four percent say that they believe in God" (GOLDSTEIN/CONNELLY 1998). DENNETT writes of religions, "They have kept *Homo sapiens* civilized enough, for long enough, for us to have learned how to reflect more systematically and accurately on our position in the universe" (1995, p518). Yet, from the record, the majority of people reflect on our position in the universe in much the same way that they did before DARWIN. The refusal of religion to die has become an embarrassment (D'AQUILI/NEWBERG 1998). I believe that part of the explanation for this lies in the nature of the mind itself.

"Okay", the reader might say; "we can agree that religious belief has been of prime importance since the beginnings of human culture. So what? What does that have to do with natural selection and other natural forces? Are you suggesting a marriage of heaven and earth, with religious belief an offspring of God *and* Mother Nature?" No. I argue that the products of imagination, including religious belief, are natural products (memes: cultural material, based on genes: DNA), and that the brain structures to conceive and store such belief are natural structures that aid human survival. However, I am suggesting a somewhat different view of *nature*: the nature of "human".

In relation to human phylogenetic processes and cultural change, LORENZ notes: "If we discover that certain behaviour patterns and norms of social conduct are found in all human beings in all cultures in exactly the same form, we can assume with virtual certainty that they are phylogenetically programmed and genetically specified" (LORENZ 1977, p182). While the content of religions differs from culture to culture, "the behavior patterns and norms" of seeking meaning and continuity in life, of searching for supersensory powers, and developing belief in such powers, this seems to be present in all existing cultures, even those practicing Buddhism (BROWN 1959; SMITH 1958). This behavior existed at the dawn of civilization some ten thousand years ago, and, I suspect, existed earlier, shortly after the advent of imagination in *Homo sapiens*. D'AQUILI and NEWBERG, based on their neurological research, make the claim that "the brain constructs gods, spirits, demons, or other personalized power sources with whom individuals can deal contractually in order to gain control over a capricious environment" (1998, p191). Religious behavior seems part of the nature of "human", in dual existence with the animal behavior (PERSINGER 1987).

As scientific investigation led to the seemingly impossible dualism of light: "how can something *really* be both wave and quanta?" so too, I suggest, has scientific investigation led (at least temporarily) to a certain dualism of the human animal: *animal* in the evolution of all its quantifiable parts, *human* in the evolution of *mind*, and in behavior based on beliefs. LORENZ goes so far as to say, "the human mind—and this one can say without exaggeration, is a new kind of life" (1977, p172). If so, a DARWINIAN approach should be to look for a new situation that might have led to "this new kind of life", which appears to be both animal and some other form that is still evolving. Human survival requires nourishment for this new form: the mind, as well as nourishment for the animal housing it. The mind and the animal brain seem to be shaped differently; they seem to have different needs. This conception is compatible with a materialist philosophy; brain and mind fit together. However, they do not simply fit together. The irregular-shaped human mind cannot be pounded into the preexisting animal brain cavity; the attempt to do so seems to be based on a preconceived belief of "mind" as merely an extension of brain ("what else could it be?"), and a yearning for evidence to support that belief. A structure for doubting reality, for imagining an alternate reality, adapts a qualitatively different approach to the world than does the animal brain (BERES 1960). It seems to have evolved to make use of that different approach for survival purposes. This does not mean that such a structure is in all respects superior for survival. LORENZ, commenting on the double edge of the reflecting process, "man's greatest discovery in the history of the human mind", states that it was "immediately followed by the greatest and gravest mistake—that of doubting the external world" (1977, p15). I suggest a simple explanation of that doubt: "man's greatest discovery" was made to do *just that*: to justify the doubting of an external world that showed human death as the final reality of life. Denial of death and other "unacceptable" realities seem an inherent part of human emotional life (BROWN 1959; BECKER 1973; LANGS 1996).

It seems plausible that an inner-directed sense, a sense "that creates images of what is not actually present", was not designed to search the external environment for food; the preexisting primate

brain, 50 million years in its formation, did that quite well. Other primate species, lacking human imagination, have survived quite well, in diverse and changing environments (MITHEN 1996). There is a counter theory in "the psychology of imagination, pointing out the origins of this capacity in the development of object consistency relative to the stimulus of vanished objects" (RANGELL 1988, p63); this is similar to BERES idea of the mental representation of an absent object (1960). Such a useful capacity, I suggest, might develop as a byproduct. It is useful to preserve an image of a real object after it vanishes from the senses. However, it would be *essential* for a created image, an "object" that could never otherwise return to be "sensed". Storing would be the only way to "sense" the image of something that never existed, and without imagination would never return—could never return. It is the preservation of such unreal and distorted images, vital for human survival, as I hope to show, that would drive the evolution of imagination. Once in place, imagination would serve in many other ways for human survival. Primate senses were old and successful devices in hunting within the real environment. An inner-directed sense, creating its own reality, would seem to have been designed for some different purpose, one that would lead to nascent psychological adaptations for survival. MITHEN, in considering evolutionary research, speaks of "integrating material from evolutionary ecology and human psychology … [towards] a DARWINIAN psychology [that] lies ahead" (MITHEN 1989, p492). I suggest that research in the evolution of human imagination should be an important part of such a DARWINIAN psychology. RANGELL describes the functional evolution of imaginative products known as fantasies: "With reference to the linkage between the cognitive and the affective, fantasies are cognitive products designed to produce a wished-for affective result. The aim is to produce pleasure and safety, while keeping anxiety or any other form of unpleasure at bay" (RANGELL 1988, p65). Just such imaginative mental activities are involved in the process of developing religious behavior and in other responses to awareness of death.

Recent work in neuroscience identifying areas of the brain especially active during religious experience (RAMACHANDRAN/BLAKESLEE 1998) and on parts of the brain's autonomic systems stimulated by religious rituals and practices (D'AQUILI/NEWBERG 1998) may advance part of my argument. Cutting across disciplines, an important area barely touched on in this paper, is gender studies: studies of the special role women must have played in the evolution of imagination and religious behavior.

Women, from earliest times seen as the source of life, would also, with stillbirths and infant deaths, be seen as a source of death. Even before imagination was directed elsewhere, women would naturally be looked to for clues in understanding the great mysterious processes of birth and death. Not only men, looking from the outside, but women themselves, as imagination evolved, would surely give special thought, imaginative thought, to their bodies and to understanding themselves in relation to life forces and those of death. Not merely as the fertility figures shown in early artifacts, women, at an early stage in human evolution, would naturally be looked to for spiritual guidance, based on their special bodily intelligence. Elaine SHOWALTER describes such special female awareness as "the corporeal ground of our intelligence" (1998, p. 338). Imagination would also find employment in sexual relations and pair bonding. Far beyond my ability to explore here, I suggest that, early in the evolution of imagination, sexual dreams and fantasies would intertwine with fears of death and hopes of some rebirth. It might well be that from this bonding would come some of the first imaginative communication: "the earliest source of the profound and complicated relation in human life among sexuality, aging, the certainty of death, and the knowledge of time" (FRASER 1988, p488). At the very least, such use of imagination by a bonding pair would tend to dispel morbid thoughts and lead to better survival strategies. As apprehension of death developed, women, in using their power and ability to choose a mate and potential father, would tend to favor one who at least offered some alternative reality mitigating the pervasive threat.

The Evolutionary Pathway to Imagination

What might have driven the engine of evolution to such a unique adaptation as that of human imagination? I suggest the following rough-and-ready account of a long complex process, as a likely sequence of events. In this account, complex questions of the nature and function of *self* and *self-awareness* will, of necessity, be simplified. In human evolution, there came a stage when big-brained, curious hominids, having practical tools but lacking those associated with mind, took an evolutionary pathway leading to human self-awareness and awareness of "other", perhaps as the inevitable consequence of their smartness and inquisitiveness.

This perception of one's individual existence in space and time as separate and in potential opposition to other human existence and the rest of nature would become a driving force in the evolution of the human animal (BRONOWSKI 1977; DOBZHANSKY 1964; LANGER 1982, 1972).

The beginnings of this new sense of identity may have taken place about 200,000 years ago, when our moribund *Homo erectus* ancestors, with brains almost as large as our own, appeared to be dying out of boredom. What awareness and what thoughts, beyond the dull archeological evidence, might such brains have expressed? "*Homo erectus* appeared about 1.5 million years ago, and survived until several hundred thousand years ago… The brain was enlarged, at first by about 20 percent, to 900 cc, but eventually, in late *Homo erectus*, to 1,100 cc, or about 80 percent of modern human cranial capacity" (DONALD 1991, p112). Such a creature would be capable of considerable self-awareness, as well as glimmerings of change and of time beyond the current moment (MITHEN 1996). Other primates may have some semblance of self-awareness, but human self-awareness, awareness of self in time, and the ability to represent such awareness, seems qualitatively different (DONALD 1991; BRONOWSKI 1977; LORENZ 1977). Being able to step back from what is represented confers a freedom to thought processes. This is crucial for the development of self concepts (DEACON 1997). DOBZHANSKY called this development "an evolutionary novelty; the biological species from which mankind descended had only rudiments of self-awareness, or, perhaps, lacked it altogether" (1967, p68). Preceding the development of human awareness of time and self "there arose, it seems, the need for a new, internally generated image, an executive agent to help coordinate action and thereby provide flexible responses to the unpredictable behavior of food, friend, and foe" (FRASER 1988, p489). Although this evolutionary process must have been gradual, we can imagine that receptive ancestor who, in bending down to drink from a still pond, stops and gazes at herself, consciously moves her head and hand, and becomes aware of her power to do so. She then comes to recognize that beyond this new-found self lies a larger world of other, no part of which can she move without first grasping it. She may also become aware of some change in herself. Much of this precocious awareness is likely to die with her. Some of it will be passed on. She will have some communication ability that allows her to struggle with and crudely express such awareness. That awareness would also serve the human need FRASER describes and this too might lead hominids further in the direction of symbolic language (DEACON 1997).

There would be an iterative process involving self awareness and language acquisition. If that process begins with awareness of self, our receptive ancestor, in achieving some glimmer of it, takes a precarious step that will lead her children's children (likely, hundreds or thousands of generations later, but quite rapidly in evolutionary terms) to nascent awareness of mortality. "Self-awareness has, however, brought in its train somber companions—fear, anxiety and death awareness" (DOBZHANSKY 1967, p68). Undoubtedly, this awareness would also advance words and grammar of human language as the way to express it, as a parallel development to the earlier form of communication: calls and gestures (DEACON 1997). Evolution is unevenly slow. Human evolution seems marked by periods of "punctuated equilibrium", by certain speciation events that cause dramatic changes (GOULD/ELDREDGE 1993). There might have been many such events and changes on the road to humanhood. Imagination, and language to express its products, would parallel those changes.

Although this analogy can take us only so far, compare the awareness of mortality that would follow an awareness of self to the awareness of cold that would follow the most recent advance of ice. With increasing awareness of cold (and a less protective outer surface than hominids had during the previous ice age) would come discomfort, pain, and eventually, some disability in hunting and in other survival tasks. Initially, nothing need have been done; individuals and tribes could suffer and survive. Nature would favor traits that increase the body's ability to withstand cold. At some point in time, with increasing cold, the more successful hominids would have gone through certain adaptations of brain and behavior, would have developed sufficient smartness and dexterity to fabricate protective covering from animal skins (DOBZHANSKY 1964). Now consider awareness of mortality. With this too would come discomfort and pain of another sort, and eventually, this might lead to some disability: some apprehensive state of mind which might reduce effectiveness in hunting and in other survival tasks. As with the cold, initially, nothing need have been done; individuals and tribes could suffer with their painful emotions and survive. Here too natural selection would be at work, favoring traits that might increase the brain's ability to withstand the painful emotions, and thus, potentially debilitating fear would be reduced and those individuals

would tend to be more fit in performing survival tasks. Fear is a powerful emotion leading to adaptations for physical protection. Fear of a kind for which physical survival precautions could not be taken would require a special kind of adaptation. I suggest that with increasing awareness, the more successful hominids would have gone through such adaptations (and have been the beneficiaries of other natural forces); they would thus develop smartness and dexterity of a different kind in order to "fabricate" the protection offered by embryonic beliefs and spiritual presences. Imagination, I suggest, would be an adaptation in that evolutionary process. Further, this would tend to drive communication along the road of symbolic language. You can't communicate religious search and discovery via calls and gestures—no matter how complex these calls and gestures might be. LANGER describes the evolution of awareness of mortality as a still incomplete process:

"With the rise and gradual conception of the 'self' as the source of personal autonomy comes, of course, the knowledge of its limit—the ultimate prospect of death. The effect of this intellectual advance is momentous. Each person's deepest emotional concern henceforth shifts to his own life, which he knows cannot be indefinitely preserved… As a naked fact, that realization is unacceptable; there are few societies, savage or civilized, that admit it today" (LANGER 1982, p103).

On that long road to awareness (which we are still traveling), hominids would become aware of changes in the environment and begin to detect what they would later come to know as life cycles. They would gain awareness of a time beyond the immediate moment. DONALD, in contrasting the awareness of time in humans to that of apes, writes:

"Their lives are lived entirely in the present, as a series of concrete episodes, and the highest element in their system of memory representation seems to be at the level of event representation. Where humans have abstract symbolic memory representations, apes are bound to the concrete situation or episode" (DONALD 1991, p149).

In order to express time beyond the immediate moment, proto-humans had to have some conduit for such thought, some linguistic structure. In the iterative struggle to communicate such thought, previously developed calls and gestures indicating food, courtship, and other opportunities; predators, storms, and other dangers, this utilitarian communication would need to be extended into more conceptual domains, and into what would eventually become symbolic language. Such shifting of communication into symbolic language would facilitate further conceptual awareness, culminating in awareness of mortality: some crude "imagining" of the possibility of one's own death. There would now be a greater need for symbolic language, and, undoubtedly, a life and death struggle to achieve it. Although other animals have complex communicative behavior, even a simple language seems impossibly difficult. This poses a profound riddle in understanding the origin of language (DEACON 1997).

A possible answer to this riddle might be found within the developing awareness of mortality and the emergence of imagination. "Imaginative discoveries" require just such a communication device as we have—language, with its set of vocabulary and grammar for communicating the abstractions of imaginative discoveries. At a later stage: "Language and awareness of personal mortality brought with them the emergence of burial practices, rituals, and symbols related to the death experience, along with the origins of religions" (LANGS 1996, p131). "Once symbolic communication became even slightly elaborated in early hominid societies, its unique representational functions and open-ended flexibility would have led to its use for innumerable purposes with equally powerful reproductive consequences" (DEACON 1997, p349).

Accidentally and reluctantly aware early humans (such awareness, I argue, emerging as an impediment to survival) would be forced to consider first the possibility, then the likelihood, and then the yet unthinkable fact of individual death. Imagination, in its early development, although it would prove to be a vital aid in dealing with awareness of death, might also have exacerbated the awareness itself by making it more vivid: "in the evolution of mind imagination is as dangerous as it is essential" (LANGER 1982, p137). Good things hardly ever come easily or without a price tag. Imagining the possibility of one's death was (and of course still is) an awesome and potentially debilitating awareness, a pervasive "danger" for the individual that cannot, as with specific threats to life, be guarded against. Survival now required something in addition to the satisfaction of physical needs: structures, devices, and processes, for the individual and then for the group, to ameliorate that difficult-to-live-with awareness. Initially, there might be little more than vague feelings of something wrong or threatening. At the very least this nascent awareness would lead to thoughts not conducive to happy hunting. How might that individual deal with such thoughts?

Consider the myriad of religious, social, and psychological support systems, the ability to commiserate, the diversions from morbid thoughts that have been developed throughout history, to mitigate and anesthetize that awareness (BECKER 1973; BROWN 1959; HOCART 1954). We have evolved, somehow, to function with a minimum of morbid thoughts, busy from day to day, planning for the future, and confident that we will live to see it. A human "requires 'understanding' not only of his world of survival, but eventually of the immaterial world of thought that is the creation of the increasing complexity and subtlety of his own process of cognition" (LAUGHLIN/MCMANUS/D'AQUILI 1990, p242). By communication and demonstration, we pass on "immaterial world of thought" survival supports to our children, as they become aware of self and death (DONALD 1991; LANGER 1982). Newly aware hominids would have lacked such supports, lacked the linguistic ability to communicate and commiserate, and hence, would have less confidence and less ability to function in survival tasks than they had before.

That you and I will die is as plain as the proverbial "nose on your face", and as difficult to look at. To avoid damaging that cognitive inner eye, we tend to look at our death much as we look at the sun: peripherally. As with the near impossible act of putting our hand into fire so as to feel the heat, we face the near impossible act of putting our thoughts into death so as to "feel" it. Indeed, we go to great lengths to avoid such activities. BECKER shows us that the entirety of human psychology is rooted in a massive attempt at denial of death. As he describes the difficulty; "the fear of death must be present behind all our normal functioning, in order for the organism to be armed toward self-preservation" (BECKER 1973, p16). A growing body of terror management theory suggests that "the most basic of all human motives is an instinctive desire for continued life, and that all more specific motives are ultimately rooted in this basic evolutionary adaptation" (PYSZCZYNSKI/GREENBERG/SOLOMON 1997, p1), and that, over thousands of years, culture has developed to manage the existential terror brought on by awareness of mortality. "But the fear of death cannot be present constantly in one's mental functioning, else the organism could not function" (BECKER 1973, p16). There are aspects of our individual death that we dare not look at and other aspects that we cannot look at. For example, try to imagine yourself dead; where would the *you* who is looking be? There seems to be no place from which this "imagining" could be emanating (FLEW 1993). We are able, through most of our lives, to put aside our individual death, to act as if *it* were not there, or as if we might change the reality of its existence in good time. Indeed, there are diverse human-devised systems that allow us to do so, elaborate systems of belief in some continuity of existence after death, and simpler systems involving ways of looking at "reality" so that "the finality of death does not exist at all" (HOCART 1954, p87). Our imagination is well designed to make use of such systems to redirect our morbid thoughts and our unpleasant sensations. We seem quite able to countermand the external senses when this suits our purpose, to alter sensory information and thus *revise* the world that would otherwise be seen, heard, and touched (KOESTLER 1964). "Mental images have the power to affect us in many of the same ways as our perceptions of reality. Consequently, our imaginations can afford us a means of experiencing loss and then being able to rejoice in still having the 'lost' object" (VICKIO 1994, p611).

In "Can the Subject Create His World?" METZGER gives an historic overview of ideas suggesting that the world, or some significant part of it, is created by mental acts. He is interested in perception rather than imagination, and he concludes that "perception is not a way of adding new facts to the world—this is the task of art and invention—but to find what there is before perceiving begins, which has not yet been found by the present perceiver" (METZGER 1974, p67). Part of what there is that has not yet been found is in the task set for imagination; perception and imagination coexist and function simultaneously. No other animal has developed such ability to willfully embellish sensed information, to perceive that which is not sensed, to deny that which is, to fantasize, to hallucinate, or to imagine things that have never existed and things that may never exist; no other animal has the linguistic ability to communicate such things. No other animal has shown the need for it; indeed, for others, these abilities would be impediments to survival. Other animals might, for survival purposes, behave so as to deceive others. Only humans seem capable of self-deceptive images, since only humans have evolved with *imagination*: reality-distorting input to the brain and nervous system. Examples of intentions in the literature of human self-deception include: avoiding pain and painful reality, evading trauma, and seeking comfortable beliefs (MARTIN 1997). All these would apply to newly aware-of-death hominids who had the nascent imagination with which to achieve such self-deception. The senses and nervous system of higher animals, including humans,

function to supply the brain with accurate information for survival. Our external senses act as hunters, perceiving as well as sensing the environment for accurate information (GIBSON 1966). At that stage of human evolution "when our early ancestors first noticed the images in a pool of water, or the shadows of things, and especially when they began to make pictures, we may fairly assume that they became puzzled about the problem of appearance and reality" (GIBSON 1966, p310). If we look at a field and "see" a fish moving across it, our control center, in effect, signals our eyes to look again; perhaps the field is a lake or pond. Or, if it remains a field, perhaps the fish is a rodent or some other land animal. (A similar example could be given for what our ears might hear.) I suggest that our internal senses developed to act as hunters of a different kind: rather than for accurate information, they hunt for that which is acceptable, and "meaningful". If they report a "senseless" presence or voice, an embellished memory, such a report might well be acceptable and even welcomed. Although "senseless", it might "make sense". It would be particularly welcomed if such a presence had been sought. But since our external and internal senses function simultaneously, it is not always clear what kind of information and experiences they (we!) are hunting for. STADDON and ZANUTTO refer to examples "of organs evolving (often not very efficiently) from one function into organs that serve a very different one" (1998, p241). In the case of the brain with its sensors, one can consider the organ to have retained its original function while taking on the added function of imagination. How might such conflicting functions of the human brain have evolved?

The Prehistoric Background

Although there is little clear evidence for how "*Homo religious* and Its Brain" evolved, (HOLMES 1996, p441), we do know quite a bit about how pre-religious brains evolved, how new structures grew around older ones, and how animals housing larger brains grew smarter in processing information from their environment. But such informational smartness, nuts-and-bolts-survival smartness, need not lead to, and certainly does not explain, the development of human imagination and such things as spiritual experiences; *they* are different kinds of *things*. Humans, together with the evolution of bigger and smarter brains, obviously useful in the struggle for survival, evolved with an addition to that system: a companion device whose usefulness is not obvious (DONALD 1991; MITHEN 1996). It is a device that gained awareness of itself and its fate, and then developed structures and processes: conceptual thought and symbolic language, to bear the weight of that awareness, to make sense of the world, to discover "meaning" in it, and hence, to make fragile, finite existence more bearable (LANGER 1982). Humans use them to ascribe "purpose" to a seemingly indifferent environment, to ameliorate frightening sensory information, to countermand unacceptable empirical evidence, and to create an altered image of the world which is then stored as mental images of what is not actually present, of what has not been actually experienced (LORENZ 1977; LANGS 1996).

Before gaining imagination, the actions of our ancestors were, undoubtedly, operant behaviors, responses to stimuli that operate on the environment. DENNETT calls such lower-order animals, creatures susceptible to operant conditioning: *Skinnerian Creatures* (1995, 1978). In time, our ancestors developed the ability to learn about the environment in ways far beyond mere trial and error behavior. DENNETT speaks of creatures like ourselves that "have *two* environments, the outer environment in which they live, and an 'inner' environment they carry around with them… we are talking of the evolution of (inner) environments to suit the organism, of environments that would have survival value in an organism" (1978, p77). DENNETT further explains: "the inner environment is simply any internal region that can affect and be affected by features of potential behavioral control systems" (p79), an environment in which advanced creatures ask, "what they should think about next" (1995, p378). Having such an inner environment, hominids, as they evolved towards human self-awareness, observing the violent end of a young comrade, the weakening and deterioration of an older one, the long sleep without an awakening, from all this, a new kind of behavioral response would begin to emerge. There would be a need for new and different survival adaptations. As awareness emerged, perhaps in dreams, first the gnawing feeling, then the shocking thought, must have taken hold: "This may happen to me". Later, the more awesome extension of thought: "Death is a common happening. This *will likely* happen to me". A less dramatic but quite likely scenario is that proto-humans came to that thought gradually over generations of increasing melancholy, moving toward depression, encountering death with a growing awareness of mortality and a feeling of helplessness. They would look at dead comrades and feel some-

thing bad, dangerous, even ominous, without knowing just what or why. They might have been prey to psychosomatic illness while in the very process of developing psyche! They might engage in unproductive searches for the cause of their feeling the physical presence of something bad or dangerous, the cause of this new kind of fear.

"Evolution has bred into the members of every animal species a rate of production of fear which corresponds to the average degree of endangerment in which the species must live and survive" (LEYHAUSEN 1973, p254). With most animals, production of fear is limited to present dangers: dangers that can be guarded against. Some animals are faced with incessant danger:

"An animal of this kind can better afford to go without food or sleep for a whole day or even longer, or to miss a mating, than to relax its constant alertness for five minutes… As long as the endogenous production of fear roughly matches actual endangerment and the overall harmony of the instinct system which has been won in the process of evolution is maintained, then for the organism concerned this is only right" (LEYHAUSEN 1973, p254).

Constant alertness could not be "right" for humans faced with nascent fears of dying, even if its initial negative effects were minimal. Human survival was already precariously balanced (MITHEN 1996). Natural selection could not provide any state-of-the-art flight or fight adaptation. Some adaptation quite new in nature was required if the species was to survive. Indeed, only one species, of those who might have had nascent awareness of mortality, perhaps the only one to develop imagination, did survive.

Of such fear and "subjective emotional experiences" LEYHAUSEN speculates: "the relationship between the propensities or instincts of fear and the experience of fear as seen from the view of the ethologist are unavoidable… in part still hypothetical and insofar represent an appeal for the development of a research program designed to test them" (1973, p255). In regard to genetic differences, from atrophy to hypertrophy of fear:

"If hypertrophy has affected the production of fear, we get the whole range from the overfearful to the serious case of anxiety neurosis, where the minimum level of the automatic production of fear has shifted considerably farther 'upward' and thus does not fall victim to atrophy from disuse even when there is a complete lack of adequate releasing situations. The person affected is therefore constantly under pressure from the strongest appetences for fear, looks for and finds a 'substitute object', and since this is, of course, not the real cause of his fear, in this instance no habituation to stimulus or decline in stimulus-specific sensitivity can set in" (LEYHAUSEN 1973, p267).

Over generations, with increasing awareness of changes and then a glimmering awareness of time itself, individuals must have struggled with increasing nonspecific fears before grasping the specific (if yet unthinkable) fear: "Death can happen to me". Such nascent feelings and thoughts may have occurred in many forms before taking root in the soil of mind as a specific fear of death. Earlier, an individual sense of life would be somewhat diffuse and impersonal: not strongly felt as a single being (LANGER 1982). Gaining self-awareness, the individual would gain an increasingly specific fear of death of self. Full awareness of a personal death is still evolving as we enter the 21st century. In the early stages of its evolution, neurological structures and language to express such thought, as well as the thoughts themselves, most likely would develop in parallel: the need and the adaptation to serve that need.

There are many aspects of this growing awareness that must be considered here. For one, human groups during the periods considered here were small: some few dozen individuals (perhaps as many as ten dozen) living together as sub-groups sharing a common space (MITHEN 1996). Bonds of friendship would tend to be strong; individuals would be mentally as well as physically important to each other. Picture now, as awareness was developing, one of a myriad of events: the death of one individual after some period of suffering, with the others trying to give aid. The dead body likely would be salient for some time before burial or other disposal. The others would struggle to come to terms with the event. Two comrades might share looks, tears, and moans; they might then, somehow, create a way of remembering and communicating their sharing the event. In time there would be a symbolic representation of the event that would be stored in nascent memory and retrieved later, around some similar event. I suggest that, in the iterative process, the development of imagination and its cultural expression would be advanced. In considering how you and I differ from our ancestors in facing death, these two matters should be considered first. We do not often look on death; we have language to share, culture and imagination with which to filter thoughts when we do look. There are comforting religious beliefs, but even for those who reject such, there are cultural supports to lean on.

Considering the controversy over the respective roles of genes and memes in evolution (BOONE/SMITH 1998), a useful analogy with the development of imagination may be that of the development of fire. There is the matter of creating the initial sparks for ignition: genetic, and then the matter of fuel to expand and keep it going: social–cultural material. With regard to my thesis here, both the initial spark of awareness of the problem of mortality, and the initial spark of imagination offering a "solution", would seem, of necessity, to be genetically based. By one or another sudden variation or more gradual change to that part of the brain beyond my knowing, imagination would begin with a genetic spark. Given that spark, the "fuel" would come from the need and social–cultural material at hand.

With a growing awareness of mortality, I suggest, would come debilitating apprehension. In order for those individuals to function and survive, that awareness, as with current mortality awareness, would need to be managed (PYSZCZYNSKI/GREENBERG/SOLOMON 1997). Such awareness would tend to be most debilitating for a creature lacking even the ability to commiserate with others: the linguistic ability to express such apprehensions. I suggest that it was such social needs, more than any direct survival need, which led and sped the evolution of those mental processes we loosely call "mind", partly individual in nature and partly communal processes: evolution of individual structures and abilities, as well as complex social organization. "Knowledge of the inevitability of death gives rise to the potential for paralyzing terror which would make goal directed behavior impossible" (PYSZCZYNSKI/GREENBERG/SOLOMON 1997, p2). How could an increasingly smart but bare-brained creature lacking cultural support come to terms with the emerging sense or feeling that he or she, as all others in the tribe, might die? FREUD in considering this question writes, "what primitive man regarded as the natural thing was the indefinite prolongation of life—immortality. The idea of death was only accepted late, and with hesitancy. Even for us it is lacking in content and has no clear connotation" (FREUD 1950, p76). The human senses were well-equipped to sense and perceive the natural world, the local environment in which to hunt and gather, to find a place to sleep secure from leopard and other predators. But how were early humans to secure themselves from this most pervasive and awesome predator? Undoubtedly, from its first glimmer, it would be a problem they would focus on. During the long hours of night, awake and in dreams, there would be few if any more vital matters of thought. How were they to avoid that sleep without end, that change of warm and vital flesh into something cold and unresponsive? Dead bodies would be salient; death itself, quite likely would be viewed as something caused by unseen predators. Nothing appeared in the sensed environment that offered a defense against these predators, nothing that the brain and its information sensors could discover. Another sensor was needed to look beyond the others, to perceive a more distant or hidden world that might offer such defenses. Needed too was the ability to share perceptions of such a world with others in the tribe, and hence, a brain with long term memory devices for storing the products of imagination.

The human brain reached 80% of its current volume about 200,000 years ago, after a 300,000-year spurt of growth (DONALD 1991; MITHEN 1996). Archaeologists can find no major change in the archeological record correlating with this second period of *Homo erectus* brain expansion; the same basic hunting and gathering lifestyle continued, with the same limited range of tools (MITHEN 1996). Thus, the first expressions of what we can identify as products of human imagination (about 70,000 years ago) occur much after the last major brain expansion. I suggest that it was in the period between 200,000 and 70,000 years ago that fear of death had reached a stage where it might have negative impact on human survival. The brain housing that fear would surely be large enough to store the products of imagination. At this stage of awareness, individuals with feelings of apprehension and an inability to deal with the perceived danger, would tend to be less fit in hunting, would be less willing to take the risks necessary for success, and would lose their leading edge in the struggle for survival. Fear involving those dangers that can be guarded against has survival value (LEYHAUSEN 1973). Fear of impending death, anxious feelings of foreboding, would tend to immobilize, and must be considered to have negative survival value. As well as individuals, entire tribes with such fear might experience higher mortality. If so, natural selection might start selecting for survival in an extraordinary way. It would not simply be those individuals and tribes who had the best physical equipment for adapting to changing environments who would prove the fittest for survival. Instead, it would be those who developed, along with physical equipment for use in hunting and gathering of food, imagination and other mental equipment for use in dampening the debilitating effects of this growing awareness and fear.

The Darwinian Dilemma

In considering natural selection and the human mind, a difficult problem for DARWINISM has been this: given that a utilitarian, unconscious brain is good and sufficient for processing sensed information and using it for survival tasks, what evolutionary pressure, what survival advantage, can be associated with sensory distortion and conscious mind? What were the stages of evolutionary transition leading to the human mind?

"The task of reconstructing the steps through which humans must have passed in their evolutionary transition is so difficult that many have chosen to ignore the problem. One extreme approach, which some may take as a counsel of complete despair, is to proclaim a discontinuity in evolution when it comes to the human mind" (DONALD 1991, p21).

DONALD goes on to elaborate the problems. "No convincing geographic or climactic conditions could have produced enough selection pressure to account for the emergence of modern humans. Hominid culture was already able to cope with a variety of climates. Although climate may have played some role, other forces must have been at work" (p209). DONALD then suggests that "the evolution of humanity is likely to have been driven at the level of cultural change" (p209). But why? "What change could have broken the constraints on mimetic culture with such a vengeance, leading to the fast-moving exchanges of information found in early human culture" (p211)? Materialists have not been able to explain this evolutionary transition. As SEARLE describes it, "materialists have a problem: once you have described all the material facts in the world, you still seem to have a lot of mental phenomena left over. Once you have described the facts about my body and my brain, for example, you still seem to have a lot of facts left over about my beliefs, desires, pains, etc" (1997, p43). At least some of these left over facts are accounted for via a God-seeking mind.

"The mind is almost as hard to define as the soul", writes JONES. As he describes the confusion within psychological theories of the mind, "it has gone from describing varieties of religious experience to censuring them, from phrenology to scanning brain and DNA, and at last—coming full circle—to explaining belief in DARWINIAN terms. Psychology is a journey from the arts to the sciences and back again" (JONES 1997, p13). On such a journey, I suggest, there is an evolutionary "bridge" to be found connecting imagination and religious behavior to the rest of adaptive behaviors. Neither anthropologists nor evolutionary psychologists have put forward a viable theory that shows why imagination and conscious distortions of sensory experience might have been more adaptive then the mindless utilitarian brain that predated them. "The brain is the ultimate lying machine" (JONES 1997, p13). Why should natural selection favor such a machine: in particular, why should it favor something that distorts reality, and hence, lies to itself? Further, nature tends to be lavish. If mind is a good survival device, why don't we find it elsewhere? Why are there no precursors of mind to be found in the rest of the animal world? (DEACON 1997; DONALD 1991; LORENZ 1977). These questions have been thorns in the side of evolutionary explanations of mind. One problem has been to explain natural selection's favoring of structures unexpressed in overt behavior: consciousness, imagination, and also, quite prevalent if not universal among early *Homo sapiens,* schizophrenia. Could schizophrenia, which (JAYNES 1976) suggests to be a vestige of ancient mind, have come into being as an adaptation for sensing spiritual guidance, and for finding a guiding spiritual voice? Looked at in terms of physical survival, these inner devices would be disadvantageous. What survival advantage could there be in fantasizing and in distorting reality? Steven PINKER suggests that we need not bother with such difficult or impossible to answer questions. He says that "we should expect to find activities of the mind that are not adaptations in the biologists' sense. But it *does* mean that our understanding of how the mind works will be woefully incomplete or downright wrong unless it meshes with our understanding of how the mind evolved" (PINKER 1997, p174).

Just so. I argue that these questions can be answered: not only how the mind works, but why. I suggest that long before discovering grain and settling in the fertile crescent to harvest it, humans had reached an evolutionary stage where "not by bread alone" was the *modus operandi*. A stage was reached where, driven by the search for supernatural support, mental considerations began to play a role in human survival, sometimes in opposition to physical considerations. Humans might, on occasion, decide to go hungry, to do (or not do) something which then resulted in hunger. They might, with the development of magic or religious belief, decide to fast, to ritualistically sacrifice food, to suffer hunger, for the sake of their mental well-being, which had come to be an important part of their total well-being and of human survival.

PINKER titles a section of *How the Mind Works*, "The Smell of Fear", in which he lists ancient and still common fears: snakes and spiders, and "large carnivores, darkness, blood, strangers, confinement, deep water … The common thread is obvious. These are the situations that put our evolutionary ancestors in danger… Fear is the emotion that motivated our ancestors to cope with the dangers they were likely to face" (PINKER 1997, p386). In this he lumps human fears with those of other animals. He does not distinguish "of mice and men", of that human apprehension put forward in Robert BURNS poem *To a Mouse*; "The present only toucheth thee; But och! I backward cast my e'e, On prospects drear! An' forward, tho' I canna see, I guess an' fear!" PINKER does not mention apprehensions: fear of future sickness or future death, fears not based on current dangers. He speaks of phobias, many of which, he suggests, we share with other animals. "The world is a dangerous place, but our ancestors could not spend their lives cowering in caves" (PINKER 1997, p388). True. But shouldn't an overview of how the mind works include human apprehensions?

Fear of eventual death, fear of dangers not in the workable environment, fears which could only be offset by imagined sources of protection, these fears only can be disadvantageous and potentially debilitating to the individuals lacking imagination. There is nothing "right" that they can do under those circumstance, but there is much they can do that is wrong for their survival. There are innate functional properties of the phenomenon of fear which evolution delivers ready made; "the individual must accept them as he must the form of his cranial bones… actively avoiding or fleeing from dangers offers the individual better prospects of survival than passivity. It does, however, also contain the possibility of doing the wrong thing" (LEYHAUSEN 1973, p250). The functional properties of human fear, of course, were and are highly complex in nature, and beyond the scope of this paper. One large topic, untouched here, is the relationship between fear and aggression (BECKER 1973; LEYHAUSEN 1973). However, it does seem reasonable to conclude that for the hominid lacking imagination, fear of an unavoidable danger, would surely increase the possibility of his doing the wrong thing—which, in the instance of a devitalizing fear, would include doing nothing in a situation that calls for action.

There have been some five million generations in the evolution of primates, increasingly aware of themselves as prey, and developing neurological structures to increase their security. Consider *Homo sapiens*, late in that stage of evolution, when, superimposed on those structures for security, there developed apprehensions, an awareness of mortality and an awareness of themselves as a kind of prey for which there seemed no way to increase security. Without the power of imagination, such awareness, I suggest, would be an impediment. Individuals encumbered with fears for which precautions could not be taken would be less successful. I suggest that an individual with such apprehension would be more hesitant in hunting big game, and less willing to take the necessary risks. The individual beginning with such fear would be less fit in making a living. The hunter who starts out hungry but apprehensive would tend to be less successful than one who starts out merely hungry. In the aggregate, entire tribes of such hungry but fearful hunters would tend to be less successful. What adaptation could be developed in response to such fear? Who would now be fittest to survive? Natural selection might favor "lesser-brained" individuals who, still secure in their ignorance, lacked awareness of impending death. Instead, a genetic spark might somehow appear, natural selection might somehow "stumble upon" a brain companion of sorts: a device or process whose function would be to hunt out, via images, ways and means of mitigating the debilitating fear.

The Language of Imagination

Symbolic language ability would undoubtedly be a necessary part of in this new kind of "hunting". Utilitarian communication, complex calls and gestures for use in hunting, most likely long predated this stage of evolution (DONALD 1991; MITHEN 1996). With awareness of mortality, the task for newly developing thought and language would be to make death livable, to formulate mitigating conceptions around death that would become the precursors to magic and religion and also to imaginative stories. These would be impossible tasks for communication systems based on calls and gestures (DEACON 1997). Via imagination, perception of the external world could be altered. Via stored images and symbolic expression of thought, apprehension of death could be ameliorated (HOCART 1954). What better way to spend countless generations of long cold nights, countless winters of discontent, than around the warmth of the communal fire, struggling to discover ways to make the newly experienced fears bearable? Perhaps there was something within the body that did not die. Per-

haps there were spirits (or demons even) who controlled such life. We can sense the struggle with such questions in the early expressions of art, the search that would lead to a variety of "answers", some of which might also be terrifying. Spirits, even demons, no matter how terrifying, would be less terrifying than death itself (HOCART 1954). As fire was used to ward off the leopard, spirits might be used to ward off death, or to provide another life. Those individuals or tribes armed with such protection against death might be more willing to take the risks necessary for a successful hunt. The spirits might indeed inspire individuals to hunt more courageously than before. Who in the five thousand years of recorded history has been more courageous in situations requiring courage than those inspired by spirits or gods?

To look at a world beyond that which is sensed, in "mythic modes", requires that aspect of imagination which in literary theory is known as "suspension of disbelief": a willingness to suppress doubt (DONALD 1991). Suspension of disbelief, I suggest, is similar to an older use of imagination necessary in order to reconfigure the world to accommodate spirits associated with the dead. The mere telling of stories would make long winter nights less monotonous, but that would hardly drive natural selection towards the adaptation of human imagination. *Homo erectus*, quite likely, lived a million years in such monotony, bored, perhaps, but genetically successful. The storytelling would need to be driven by matters of life and death. DONALD shows life and death mythic constructs to be among the oldest of human inventions:

"Even in the most primitive human societies, where technology has remained essentially unchanged for tens of thousands of years, there are always myths of creation and death and stories that serve to encapsulate tribally held ideas of origin and world structure… These uses were not late developments, after language had proven itself in concrete practical applications; they were among the first" (DONALD 1991, p213).

In discussing the prime uses of language, DONALD adds: "Initially, it was used to construct conceptual models of the human universe. Its function was evidently tied to the development of integrative thought—to the grand unifying synthesis of formerly disconnected, time-bound snippets of information" (p215). To integrate and express life and death thoughts requires that language we now associate with imagination and mind activity. Such thought and such use of language would, I suggest, from its beginning, intertwine with utilitarian communication, and with the older form of calls and gestures. Once in place, imaginative mind processes would function alongside those of utilitarian brain in human communication, along a continuum from purely sensory expressions to those that are inner directed and conceptual. We see such in current communication, in a continuum from work-related statements, questions and commands, where accuracy is required, to those in religion and poetry, where ambiguity is acceptable and even encouraged. Also intertwined with such expressions of thought in human communication are certain "pseudo-symbolic structures… emotions, feelings, desires. They are not symbols for thought, but symptoms of the inner life, like tears and laughter, crooning, or profanity", (LANGER 1957, p83). These structures too, I suggest, would evolve alongside the emerging human mind, to express the fear, the apprehension, and later, the joy and other good feelings involved in the new search and discovery process.

Consider that era in prehistoric time when awareness of self and of death-of-self first emerged and found expression. Before this time, the essential role of language would be to communicate as accurately as possible: danger and opportunity, sighting of a predator, sounds of an antelope herd, where food was to be found, when and how to secure it, who should perform the various tasks involved. Plain, concrete, unambiguous communication was needed for success. The payoff was meat or plants that safely could be eaten. With hunting–gathering of information relating to the dead, with tasks related to spiritual well-being, the role of language and pseudo-symbolic structures would be to communicate these thoughts and emotions: death-mitigating ideas and fears, in such a way that belief systems could be built. The payoff was an effective spirit or a god that could be believed in.

Natural selection would favor individuals who "successfully" came to terms with death: who used their emerging minds to find ways of making death bearable. Imagine that stage in evolution when *Homo sapiens* first became aware of the frightening mystery of non-accidental death, of fatal illness, of an aging process toward certain death. Lacking knowledge of disease, they might have feared that death itself might be contagious (LANGER 1957). From their own terrible dreams they might have looked at a dead comrade and wondered as Hamlet wondered; "in that sleep of death what dreams may come?"

"To the dreamer dreams can be just as real, just as rich in experience. Is the world perhaps only a dream? Thoughts such as these must have struck with overwhelming force the man who had just emerged from the twilight of an unreflective, 'animal' realism, and it is understandable that, beset by such doubts, he should turn his back on the external world and concentrate his whole attention on the newly discovered inner world" (LORENZ 1977, p15).

The origin of spiritual belief in connection with death is to detach the survivors' memories and hopes from the dead (FREUD 1950). What "tools" might be found or made to "reshape" death? Those who had the ability to think such questions were on their way to answers. Early humans who developed rituals to mourn the dead, and then developed magic or religion to make the apprehension of death "bearable", would function better: would tend to be less debilitated by fear of death, as individuals and in community. The foundation of all ritual is that one cannot do it alone. The individual cannot impart life to himself; others: human or superhuman, are needed (HOCART 1954, BECKER 1975).

Natural selection might have favored altruism: such behavior might have had an evolutionary component—favoring those tribes, as well as those individuals within the tribe, who demonstrate altruistic behavior. "There is ample evidence that humans cooperate with people to whom they are not closely related—more so than for any other species… Humans, however, have evolved dispositions to cooperate or compete that take their cues from the actions of other individuals" (SULLOWAY 1998 p38). In a similar way, natural selection might then favor tribes as well as individual members who, having come to the realization that everyone, including themselves, dies, developed the ability to make death bearable—for the tribe as well as for themselves as individuals. Natural selection would then favor those with the ability to imagine and explain, to create and socially share ways of making death bearable. Ultimately, they would search for and find spirits and gods. Memory: stored imagination, would now get to be a communal process, a unique social process for preserving "the meat" hunted down by individual imaginations. Our ancestors, hunting for game with their appetites set on finding antelope meat, might have to settle for lesser game, or even for vegetation that merely took the edge off their hunger. These ancestors, hunting for spirits with their minds set on finding one that could awaken the dead, might have to settle for a lesser god, or even for vaguely sensed spirits that merely offered hope.

Conclusion

All societies, in their rituals and beliefs, have transcended the reality of what their senses and experiences reveal about human death (BECKER 1973; BROWN 1959; HOCART 1954). This is true even of societies where the people deny that such death has occurred (LANGER 1957; 1982). It is also true of Buddhists, who have no God or belief in afterlife. Buddha left these matters sufficiently equivocal to allow beliefs that transcend the sensed reality and even those that "abolish" death (HOCART 1954; SMITH 1958).

The essential difference between human brains and those of other animals, the difference which I believe led to other differences, lies in imagination: an adaptation which enabled humans to wrestle with the one set of problems which no other animal has had: a problem originating with human awareness of self, and then, some shrouded awareness of impending death-of-self, and finally, the problem of how to make that awesome awareness bearable. Early *Homo sapiens*, to the extent they lacked imagination and culture built on imagination, would also lack the individual and collective support systems we now have in place to make such awareness bearable. Modern minds, drawing from past cultures, have developed abilities to keep conscious thoughts of mortality separate from day-to-day business thoughts. Thus we have learned to live and function in pockets of immortality (MONTELL 1999, 2001). We go to work each morning, wrapping ourselves in a mantle of immortality, the fabric of which is sewn in a series of plans and activities we "know" will be executed; we will not die today; we have no thought of it. Intellectually, yes: we are aware of possible mishap. Practically, no: we have developed mechanisms and processes that allow us to function day to day, week to week, and beyond, as if we were immortal. Early *Homo sapiens*, newly aware of their mortality and fearful, lacking such mechanisms and processes, would expend precious energy in a state of unproductive alertness and anxiety, and would function less well in an already precariously balanced existence.

Nature would provide the mechanisms and processes of imagination. Nurture of the human spirit would lead to the rest: untold years of development, recorded over the past five thousand years. Beyond our brief individual struggles, living under the edge of awareness of mortality, we've had long years as a species, surviving and even flourishing under this sword of Damocles nature has set for us. Intention-

alists might suggest it to be a two-edged sword: that essentially remorseless nature also has expressed some other quality by giving *Homo sapiens* the edge of imagination with which to shape religious behavior and perceptual realms of immortality—and to reshape "the self".

For perhaps the first time in human history, there is now a significant community of materialists who are facing the hard empirical evidence with regard to human life and death, without imaginative extensions of that evidence. The potential impact of this reversal of imaginative thought with regard to "the self" has barely been touched on in public discussion. EINSTEIN felt that "*the true value of a human being* is determined primarily by the measure and the sense in which he has attained liberation from the self" (1954, p12). However, this value system remains largely unexplored. This expression of EINSTEIN's mind must be viewed against the backdrop of some five billion minds that, in some form, are chained to traditional religious beliefs and to the ancient self of which EINSTEIN speaks. Each of these minds has a survival need.

Human survival is, in and of itself, a dual affair. There is all that we do that, in form, is just as any other animal does in making a living. There is also and-not-by-bread-alone behavior, survival behavior that distinguishes us from all other animals. There are psychological states, apparently unknown to other animals, in which life seems impossible or not worth living. In such states, although the body may be healthy, humans die: the mind dies, or the self commits self-slaughter—well named since it is only for the aware self that life has become impossible. The animal part (if only the self could be severed) could—and sometimes does—survive. These psychological states are imaginative states but they are as vital as the bodily states. If one accepts the logic of this dualism: animal survival and not-by-bread-alone survival, then science, in its search for human origins, must continually look beyond stone tools, economic forms, and other evidence of smart brains engaged in making a living, to the imaginative aspects of human presence, difficult though these may be to detect with hard evidence. I suggest that these imaginative aspects evolved to engage in a unique struggle based on unique awareness humans had—and have—of their environment. We are witnessing the current dynamics of that struggle.

Much of my argument here is conjecture, with some of it beyond the possibilities of unearthing hard evidence. For that, I appeal to the reader's mind to join mine in this exploration of the roots of imagination. I hope to encourage, in the biological and behavioral sciences, further investigation of the role of imagination.

Author's address

Conrad Montell, Diablo Valley College, 3150 Crow Canyon Place, San Ramon, CA, 94583, USA. Email: cmontell@attbi.com

References

Becker, E. (1973) The denial of death. The Free Press: New York.

Becker, E. (1975) Escape from evil. The Free Press: New York.

Beres, D. (1960) Perception, imagination, and reality. The International Journal of Psycho-Analysis 41:327–334.

Boone, J. L./Smith, E. A. (1998) Is it evolution yet? a critique of evolutionary archaeology. Current Anthropology 39:141–173.

Bronowski, J. (1977) A sense of the future. The MIT Press: Cambridge MA.

Brown, N. O. (1959) Life against death: The psychological meaning of history. Vintage Books: New York.

Chaplin, J. P. (1985) Dictionary of psychology. Dell Publishing: New York.

d'Aquili, E. G./Newberg, A. B. (1998) The neuropsychological basis of religion, or why God won't go away. Zygon: Journal of Religion and Science 33:187–201.

Deacon, T. W. (1997) The symbolic species: The co-evolution of language and the brain. W. W. Norton & Company: New York.

Dennett, D. C. (1978) Brainstorms. Bradford Books: Cambridge MA.

Dennett, D. C. (1995) Darwin's dangerous idea: Evolution and the meanings of life. Touchstone: New York.

Dobzhansky, T. (1964) Heredity and the nature of man. The New American Library: New York.

Dobzhansky, T. (1967) The biology of ultimate concern. The New American Library: New York.

Donald, M. (1991) Origins of the modern mind: Three stages in the evolution of culture and cognition. Harvard University Press: Cambridge MA.

Drever, J. (1964) A dictionary of psychology. Penguin Books: Baltimore.

Eccles, J. C. (1989) Evolution of the brain: Creation of the self. Routledge: London.

Einstein, A. (1954) Ideas and opinions. Crown Publishers: New York.

Flew, A. (1993) Atheistic humanism. Prometheus Books: Buffalo.

Fraser, J. T. (1988) Time: Removing the degeneracies. In: Blum, H. P./Kramer, Y./Richards, A. K./Richards, A. D. (eds) Fantasy, myth, and reality. International University Press: Madison CN, pp. 481–501.

Freud, S. (1950) Totem and taboo. W. W. Norton: New York.

Gibson, J. J. (1966) The senses considered as perceptual systems. Houghton Mifflin: Boston.
Goldstein, L./Connelly, M. (1998) "Teen-age poll finds support for tradition". The New York Times. April 30, p. 1.
Gould, S. J./Eldredge, N. (1993) Punctuated equilibrium comes of age. Nature 366:223–227.
Hocart, A. M. (1954) Social origins. Watts & Company: London.
Holman, H. C./Harmon, W. (1992) A handbook to literature. Macmillan Publishing Company: New York.
Holmes, R. (1996) Homo religious and its brain: Reality, imagination, and the future of nature. Zygon: Journal of Religion and Science 31:441–455.
James, W. (1936) The varieties of religious experience. Random House: New York.
Jaynes, J. (1976) The origins of consciousness in the breakdown of the bicameral mind. Houghton Mifflin: Boston.
Jones, S. (1997) The set within the skull. The New York Review of Books 44(17):13–16.
Koestler, A. (1964) The act of creation. Dell Publishing: New York.
Langer, S. K. (1957) Philosophy in a new key. Harvard University Press: Cambridge MA.
Langer, S. K. (1967) Mind: An essay on human feeling (I). The Johns Hopkins University Press: Baltimore.
Langer, S. K. (1972) Mind: An essay on human feeling (II). The Johns Hopkins University Press: Baltimore.
Langer, S. K. (1982) Mind: An essay on human feeling (III). The Johns Hopkins University Press: Baltimore.
Langs, R. (1996) The evolution of the emotion-processing mind. Karnac Books: London.
Laughlin,C. D./Mcmanus,J./d'Aquili,E. G. (1990) Brain, symbol & experience: Toward a neurophenomenology of human consciousness. Shambhala Publications: Boston.
Leyhausen, P. (1973) On the natural history of fear. In: Lorenz, K./Leyhausen, P. (eds) Motivation of human and animal behavior: An ethological view. Van Nostrand Reinhold: New York, pp. 248–71.
Lorenz, K. (1977) Behind the mirror: A search for a natural history of human knowledge. Harcourt Brace Jovanovich: New York.
Martin, M. W. (1985) Self-deception and self-understanding. University Press of Kansas: Lawrence KS.
Martin, M. W. (1997) Self-deceiving Intentions. Behavioral and brain sciences 20:122–123.
Metzger, W. (1974) Can the Subject Create His World? In: Pick, H. L. Jr. (ed) Perception: Essays in honor of James J. Gibson. Cornell University Press: London. pp. 57–71.
Mithen, S. (1989) Evolutionary theory and post-processual archaeology. Antiquity 63:483–94.
Mithen, S. (1996) The prehistory of the mind: The cognitive origins of art, religion and science. Thames and Hudson: London.
Malhotra, A. K. (1997) Sartre and Samkhya-Yoga on self. In: Allen, D. (ed) Culture and self. Westview Press: Boulder CO, pp. 111–128.
Montell, C. (1999) Creative imagination: Evolutionary theory's recalcitrant problem child. Psychological Inquiry 10:342–343.
Montell, C. (2001) Speculations on a privileged state of cognitive dissonance. Journal for the theory of social behaviour 31:119–137.
Parrinder, G. (1984) World religions from ancient history to the present. Facts On File Publications: New York.
Persinger, M. A. (1987) Neuropsychological bases of god beliefs. Praeger: New York.
Philippe, J. (1903) L'image mentale (evolution et dissolution). Alcan: Paris.
Pinker, S. (1997) How the Mind Works. W. W. Norton: New York.
Pyszczynski, T./Greenberg J./Solomon S. (1997) Why do we need what we need? A terror management perspective on the roots of human social motivation. Psychological inquiry 8:1–20.
Ramachandran, V. S./Blakeslee, S. (1998) Phantoms in the brain: Probing the mysteries of the human mind. William Morrow: New York.
Rangell, L. (1988) Roots and Derivatives of Unconscious Fantasy. In: Blum, H. P./Kramer, Y./Richards, A. K./Richards, A. D. (eds) Fantasy, myth, and reality. International University Press: Madison CN, pp. 61–78.
Searle, J. R. (1997) Consciousness & the philosophers. The New York Review of Books 44(4):43–50.
Showalter, E. (1988) Feminist Criticism in the Wilderness. In: Lodge, D. (ed) Modern criticism and theory. Longman: London, pp. 330–353.
Smith, H. (1958) The religions of man. Harper & Row: New York.
Staddon, J. E. R./Zanutto, B. S. (1998) In praise of parsimony. In: Wynne, C. D. L./Staddon, J. E. R. (eds) Models of action: Mechanisms for adaptive behavior. Lawrence Erlbaum Associates: Mahwah NJ, pp. 239–268.
Stephen, M. (1989) Self, the sacred other, and autonomous imagination. In: Herdt, G./Stephen, M. (eds) The religious imagination in New Guinea. Rutgers University Press: New Brunswick, pp. 41–64.
Sulloway, F. J. (1998) Darwinian Virtues. The New York Review of Books 45(6):34–40.
Vickio, C. J. (1994) Lost and found: Finding value in life through imagining loss. Death studies 18:609–619.
Webster's New World College Dictionary (1996) Simon & Schuster: New York.
Weigert, A. J. (1988) To be or not: Self and authenticity, identity and ambivalence. In: Lapsley, D. K./Power, F. C. (eds) Self, ego, and identity. Springer-Verlag: New York, pp. 263–281.

Zusammenfassungen der Artikel in deutscher Sprache

Robert Artigiani
Die Evolution des Menschen und die Evolution des Menschlichen

Die „Evolution des Menschen“ wird seitens der Biologie präzise nachgezeichnet und hat zu einem fundierten Verständnis der menschlichen Herkunft geführt. Methodologisch wird dabei im Bereich der Naturwissenschaften operiert und eine der zentralen Fragestellungen besteht darin, ob mit denselben naturwissenschaftlichen Instrumentarien und Strategien auch die „Evolution des Menschlichen“, d.h. die Evolution von Moral, eines „Selbst“ und von Bewußtsein adäquat dargestellt werden kann.

In diesem Artikel wird die These entwickelt, daß die „Evolution des Menschlichen“ eine eigene Systemebene darstellt, die nur mit einer ihr entsprechenden Sprache und Methodologie zur Darstellung gelangen kann. Als Kernkonzept für diesen Ansatz dient die „Selbstorganisation komplexer Systeme“. Davon ausgehend wird auch ein Brückenschlag zwischen biologischen und kulturwissenschaftlichen Ansätzen unternommen. Dabei wird deutlich, dass die grundlegenden biologischen Muster beibehalten werden, während die konkreten Inhalte mit denen diese Muster operieren wechseln.

Fivaz-Depeursinge
Emotion und Kognition im ersten Lebensjahr
Über Triangulierung zwischen Kind, Mutter und Vater

Schwerpunkt der hier präsentierten Untersuchung ist die sog. Triangulation, d.h. die sog. drei Personen-Interaktion die zwischen dem Kind, der Mutter und dem Vater stattfindet. Dieser Ansatz ist insoferne neu, als bisherige entwicklungspsychologische Forschung (vor allem aus dem Bereich der Psychoanalyse und der Bindungsforschung) sich überwiegend auf die dyadische Form der Mutter–Kind Beziehung konzentrierten und darüber hinausgehende Interaktionsformen nicht berücksichtigten. Als methodische Grundlage diente das sog. „Lausanner Trilogiespiel“. Im Rahmen dieses Spieles waren vor allem vier Konstellationen von besonderem Interesse: Mutter spielt mit Kind, Vater ist an der Peripherie (zwei plus eins); Vater spielt mit Kind, Mutter an der Peripherie (zwei plus eins); alle drei spielen gemeinsam (drei gemeinsam); Vater und Mutter interagieren, Baby ist an der Peripherie (zwei plus eins).

Das wesentliche in diesem Trilogiespiele erstrebte Ziel sind gemeinsam erfahrene, positive Emotionen, die vor allem im Bereich der Körpersprache und mimischer Expression untersucht werden.

In der vorliegenden Arbeit werden vor allem Daten bezüglich der sog. „sekundären Intersubjektivität“ (tritt gegen Ende des ersten Lebensjahres auf) präsentiert. Dabei zeigt sich, dass Kinder Repräsentationen ihrer Bezugsfelder auf der Basis der affektiven Erfahrungen konstruieren. Darüber hinaus werden auch Daten aus dem Bereich der sog. „primären Subjektivität“ (Alter der Versuchsgruppe zwischen drei und vier Monaten) vorgestellt.

Die Ergebnisse machen deutlich, dass bereits in den ersten Lebensmonaten Affekte und Kognitionen engstens mitsammen verbunden sind und die jeweiligen sozialen und physischen Umgebungsbedingungen starken Einfluss auf die Affekte und Kognitionen ausüben.

Sulamith Kreitler
Bewußtsein und Bewußtseinszustände
Eine evolutionäre Perspektive

In diesem Artikel soll ausgehend vom sog. „Bedeutungssystem“ ein neuer Ansatz im Rahmen der Bewußtseinsforschung präsentiert werden. Ausgangspunkt ist die psychosemantische Konzeption von Kognition wonach im Bereich der Bedeutung der zentrale Inhalt und die zentrale Funktion von Kognition festzumachen ist. Bedeutung setzt sich dabei aus Bedeutungseinheiten zusammen die zwei Komponenten enthalten: Der „Referent“, d.h. der Input

bzw. der Reiz, dem Bedeutung zugesprochen wird und der „Bedeutungswert" in welchem der Bedeutungsgehalt des Referenten zum Ausdruck kommt. Jede Bedeutungseinheit kann anhand von fünf Variablen dargestellt werden, die zusammen ein sog. Bedeutungssystem konstituieren. Jeder Stimulus wird dabei innerhalb dieses Bedeutungssystems analysiert und es werden dabei entsprechende Bedeutungszuschreibungen vorgenommen. Die Funktion derartiger Bedeutungszuschreibungen innerhalb des Bedeutungssystems besteht in der Identifikation des Input, in der Aufrechterhaltung kognitiver Funktionen, diverser Persönlichkeitsmerkmale und Emotionen.

Veränderungen von Bewußtseinszuständen werden als Folge von veränderten Bedeutungsdimensionen interpretiert, wobei im Zusammenhang damit das gesamte kognitive System (inkl. Verhalten, Emotionen, Persönlichkeit, „Selbst") verändert wird.

Diese Überlegungen werden schließlich noch um die evolutionäre Dimension der Entwicklung von Bedeutung erweitert.

Adolf Heschl

Gene für Lernen

Lernprozesse als Expression präexistenter genetischer Information

Es wird allgemein angenommen, dass Lernen sowohl bei Tieren wie bei Menschen eine besondere, da grundsätzlich nicht-genetische Methode darstellt, um neuartige Informationen aus der Umwelt aufzunehmen. Tatsächlich wird Lernen bis auf den heutigen Tag als die wichtigste Ausnahme vom genetischen Informationsparadigma verstanden. Diese Ansicht wird sogar von vielen Theoretikern vertreten, die zugleich wiederholt die Wichtigkeit des genetischen Anpassungsprozesses betonen, indem sie sich auf die allgemeine Gültigkeit des sogenannten zentralen Dogmas der Molekularbiologie berufen, welches jegliche gerichtete Instruktion des Genoms durch phänotypische Einflüsse verbietet. Wenn man allerdings einen genaueren Blick darauf wirft, was tatsächlich in einer Reihe von ganz unterschiedlichen Lernprozessen geschieht, so entdeckt man sehr schnell, dass diese dem zentralen Dogma voll gehorchen und, damit einhergehend, dem Mutationsprinzip der Evolutionstheorie, das zufallsartige Variation als die einzige Quelle evolutionärer Veränderung, das heisst: von Informationsgewinn vorschreibt. Folglich ist Lernen verstanden als echter Zuwachs an Information in dieser Perspektive nur möglich über konkrete genetische Veränderungen innerhalb der Keimbahn, das heisst durch nichtsomatische Mutationen, und hat daher auch nichts zu tun mit unserem Allgemeinverständnis von „Lernen" als einem ontogenetischen Phänomen. In der Zwischenzeit ist es bereits möglich geworden, zumindest die groben Umrisse einer solchen rein genetischen und, damit einhergehend, erstmals wirklich phylogenetischen Theorie des Lernens zu skizzieren, gestützt durch die erst kürzlichen empirischen Fortschritte auf dem schnell wachsenden Gebiet der molekularen Lernforschung.

Gérard Weisbuch

Der Ansatz komplexer adaptiver Systeme in seiner Anwendung auf die Biologie

Ziel dieses Beitrages ist die Darstellung der Anwendung von Konzepten und Methoden die aus der statistischen Physik chaotischer Systeme sowie der nicht-linearen Dynamik kommen auf bestimmte Bereiche der theoretischen Biologie. Zentraler Gesichtspunkt ist dabei die Untersuchung der funktionellen Organisation von sog. „mehr Komponenten Systemen", welche auf der vereinfachten Beschreibung individueller Komponenten aufbaut.

Im ersten Teil werden einige Beispiele komplexer Systeme aus Biologie und Physik erörtert. Darüber hinaus werden drei basale Formalismen, die in der theoretischen Biologie Anwendung finden, dargestellt.

Im folgenden Abschnitt über Netzwerke wird das Konzept des Attraktors vorgestellt und darüber hinaus die Unterschiede zwischen organisiertem und chaotischem Verhalten thematisiert. Konkretisiert werden diese Überlegungen anhand der funktionalen Organisation von Gedächtnisleistungen des Immunsystems und des Nervensystems.

Carlos Stegmann

Der menschliche Verhaltensinstinkt

Das Zustandekommen moralischer Entscheidungen

Der vorliegende Artikel beschäftigt sich mit dem Zustandekommen menschlicher Verhaltensentscheidungen, sowie den diesen Entscheidungen zu-

grunde liegenden biologischen Prozessen. Der dabei entwickelte interdisziplinäre Ansatz stützt sich auf Untersuchungen der Verhaltenswissenschaften, der Neurophysiologie und bezieht auch Überlegungen aus den sozialen und politischen Wissenschaften ein.

Die Arbeit beginnt mit der Darstellung einiger grundlegender Postulate und Begriffsbestimmungen. Am wichtigsten ist dabei der Begriff der biologischen Einheiten, der besagt, dass jede Tierart eine biologische Einheit ist, die Schichten enthält. Die Gruppe bzw. das zugehörige Gruppenverhalten stellen die oberste Schichte dar. Darunter liegen die Individuen und das zugehörige Individualverhalten. Eine noch tiefere Schicht ist der neurophysiologische Apparat des individuellen Verhaltens. Ereignisse oder Phänomene die sich innerhalb dieser Einheiten vollziehen betreffen dabei alle Schichten. Dies bedeutet auch, dass sich alle Schichten zusammen entwickeln.

Die Reduktion von Gruppen- auf Individualverhalten gelang hauptsächlich durch die „kin selection" Theorie von William HAMILTON. Diese Theorie, aus der sich der Begriff der „inclusive fitness" ergab, wurde von Edward WILSON zur Wissenschaft der Soziobiologie erweitert. In seiner Anwendung der Theorie auf den Menschen weist WILSON nach, dass viele Veranlagungen des menschlichen Sozialverhaltens genetisch bestimmt sind. Die Anwendung des „soziobiologischen Paradigmas" auf den Menschen scheitert jedoch am anscheinend unüberwindbaren Hindernis des menschlichen freien Willens. „Freier Wille" ist jedoch ein rein *analytischer* Begriff der keinerlei *synthetische* Beziehung zu Vorgängen in der Realität hat. Was nottut ist eine These die den Vorgang beschreibt, wodurch die *Information* zustande kommt, die das Verhalten des Menschen leitet und seine Entscheidungen bestimmt. Dies ist die These der Glaubensprägung, wodurch *objektive* (zweckmässige, KANTs hypothetischen Imperativen entsprechende) und *moralische* (zweck-ungebundene, KANTs kategorischem Imperativ entsprechende) instrumentale und terminale Werte im heranwachsenden Menschen *eingeprägt* werden. Die Prägung erfolgt durch die Einbeziehung der Individuen in soziale Gruppen von Artgenossen und bestimmt das solidarische Verhalten der Individuen dieser Gruppen auf der Ebene (Schicht) der VON HAYEK beschriebenen „ausgedehnten Ordnung" der menschlichen Gesellschaft. Dieses Prägungsphänomen wurde von Konrad LORENZ in seinem Werk „Die Rückseite des Spiegels" vorgeschlagen und entspricht Judith HARRIS' „group socialization" These.

Im Feld der *objektiven* (zweckgebundenen) Kognition entsteht Information durch Vorhersage und Kritik. Im Feld der *moralischen* (zweckungebundenen) Kognition hingegen, sind Vorhersage und Kritik durch den naturalistischen Fehlschluss unterbunden, und Information kann nur *ex-post-facto* durch die Versuch-und-Irrtum Methode der Evolution entstehen. Die vorgeschlagene Erklärung des menschlichen Verhaltens besagt daher, dass durch phylogenetische Adaptation die Prägung der objektiven Werte *reversibel*, die der moralischen Werte *irreversibel* ist.

Die irreversible Prägung bestimmt die *ontomoralische Starrheit* des Glaubens an die Richtigkeit von moralischen instrumentalen und terminalen Werten, und diese Starrheit ist ihrerseits grundlegend für die *Evolution* dieser Werte in der menschlichen Gesellschaft. Es entsteht hierdurch ein Erklärungsprinzip für Sozialverhalten sowie für Institutionen.

Der Artikel bezieht sich auch auf die informationstheoretischen Grundlagen der Ordnung der Lebewesen und der menschlichen Gesellschaft, und enthält eine kurze Besprechung von Werken repräsentativer Autoren die sich mit Moral befasst haben.

Zuletzt bezieht sich der Artikel auf die neurophysiologischen Grundlagen des Vorgangs der Glaubensprägung. Antonio DAMASIO hat das Vorhandensein eines neuralen Mechanismus im menschlichen Gehirn festgestellt, den er als „somatic markers" bezeichnet hat. Diese „markers" erzeugen bestimmte körperliche Zustände, die wir als Gefühle wahrnehmen und die mit wechselseitig sich entsprechenden objektiv bekannten Werten verbunden sind. Dies sind die Gefühle, die „gelernt", das heißt, in den physiologischen Apparat „Gewissen" eingeprägt werden müssen. Aufgrund dieser Untersuchungen wird jener *Entscheidungsprozess* beschrieben, der zu den konkreten Verhaltensentscheidungen führt.

Conrad Montell

Über die Evolution des Gott suchenden Geistes

Eine Untersuchung über die Frage warum natürliche Selektion Imagination und Verformung sensorischer Eindrücke begünstigt

Die Entstehung des Selbstbewußtseins und das damit verbundene Wissen um die eigene Sterblichkeit stellen zentrale Faktoren hinsichtlich der Entwicklung imaginativer Fähigkeiten dar. Als Produkte von Imagination entstehen diverse religiöse und mythische

Vorstellungen, welche hinsichtlich der Bewältigung der Ängste die mit dem Wissen um die eigene Sterblichkeit einhergehen eine zentrale Rolle spielen. Eines der zentralen Kennzeichen von Imagination ist die zunehmende Distanz zum Bereich konkreter sensorischer Erfahrungen. Imagination erweist sich aus dieser Perspektive als ein Mensch und Tier unterscheidendes Kriterium, welches alle Lebensbereiche durchdringt und in der Untersuchung der Ursprünge des Menschen berücksichtigt werden sollte.

Die Evolution imaginativer Fähigkeiten sowie des damit verbundenen religiösen Verhaltens (im weitesten Sinne) wird als adaptiver Prozeß interpretiert, der es Menschen ermöglicht Umgehensformen mit dem Bewußtsein der eigenen Sterblichkeit zu entwickeln.