Views and Reviews

Limits of Natural Science: Brain Research and Computers

Hennig Stieve

Institut für Biologie II, RWTH Aachen, Bundesrepublik Deutschland

Z. Naturforsch. **50c,** 317–336 (1995); received November 25, 1994

Epistemology, Consciousness, Language, Learning, Creativity

The criterion that pure natural science can only investigate objective phenomena which can be observed by independent observers sets certain limits to our scientific understanding of brain functions. The methods and the present state of brain research and of computer development are described. The limitations of brain research are discussed by comparing the properties of brains and computers. At least for the time being we do not know of any natural scientific – *i.e.* physical or chemical – method which allows the objective measurement of consciousness, sensations, and emotions.

Foreword

Epistemology (Erkenntnistheorie) of natural science is an area which seems to be neglected by many scientists and philosophers (Mohr, 1981). Scientific research is making such progress that most scientists pragmatically confine themselves to more promising tasks, applying science's strict methods. It is not popular among natural scientists to discuss the epistemic foundations of science and the possible limits of scientific understanding. On the other hand most philosophers also avoid this field, maybe because they do not find this area enough of a challenge or think it is too near the province of "mechanics".

One field that goes beyond the limits of natural science is brain research. In recent years, scientific brain research has made significant progress, and so has computer development – possibly even more so. Both fields have intensely stimulated each other. A number of important questions of brain research was raised anew because of this. I would like to discuss the limitations of brain research by comparing the properties of brains and computers. Are the observable capabilities of brains and computers fundamentally different?

For about two years I have been discussing this subject with brain and computer researchers, psychologists, and scholars of the humanities. These discussions produced some arguments and misunderstandings. This article is meant to further this interdisciplinary discussion.

Introduction

At the outset I want to remark that I use the word "scientific" here strictly in the sense of „naturwissenschaftlich" *i.e.* "pertaining to the natural sciences", not just „wissenschaftlich", conforming to the rules of academic discourse. Moreover when I use the word physics, it is meant to include chemistry and all the so-called exact natural sciences as well. Natural science is not the only form of scholarly research (Wissenschaft). There are several others, such as mathematics, philosophy, and other humanities, each of them characterized by methods and/or objects. Psychology and sociology, for example, apply methods both of natural science and of humanities.

Our brain has the amazing capability to understand nature

To understand nature means that we can have or make theories of the world which describe the functional relations, much of these in mental pictures which we can visualize. However, there is no evidence that our brain is able to understand nature completely, to understand the entire world. It may well be that the brain provides no understanding of certain properties of nature, especially of those which had no direct significant impact in our evolution. But since we do not know the limits

Reprint requests to Prof. Dr. H. Stieve.

of our understanding of the world, we must attempt as much as we find possible.

There are two different aspects of human brain activity:

1. The brain can be described as an information processing organ which controls bodily functions. These brain functions can be observed and measured using the methods of physics (so-called third person perspective).

2. We know the human brain to be the "seat" of psychic experiences (consciousness, sensations, emotions) which – at least for the time being – can not be measured with methods of physics (first person perspective, Innenperspektive).

What do we know about the world and about ourselves as humans?

(Epistemological Foundations of Science; Erkenntnistheoretische Grundlagen der Naturwissenschaft)

Everything which we register of the world, we experience exclusively by means of our sense organs. We cannot experience the world directly; each of us only directly experiences sensations (Empfindungen), his own consciousness (Bewußtsein). From our senses we only obtain sensory impressions which we interpret as properties of our environment (Umwelt). In our consciousness we notionally summarize certain combinations of sensory impressions (Sinneseindrücke) as things or objects. But at best our senses show us more or less adulterated properties of our environment. The "thing-in-itself" („Ding an sich" – Kant, 1724–1804) – if it exists – is not accessible to our perception (Erkenntnis).

We cannot determine whether our senses show us the reality or simulate a falsified or hallucinated environment. Actually, there is no way to find out whether the humans and objects around me – which I perceive – are real, or whether they are – in total or in part – deceptions of my senses. This holds for all properties of the world, also if characterized by physical or chemical measurements, because ultimately these too are accessible to me only by means of my senses. We cannot prove that there exists a real world outside of us. The existence of the real world is not provable. In its extreme this line of reasoning may lead to solipsism, *i.e.* the theory that nothing is real except the contents of my own consciousness (Bewußtseinsinhalte). Naturally that appears absurd to us, because the fact that the real world cannot be proven neither means that it does not exist, nor that we cannot find out anything about it. According to the theory of evolutionary epistemology (see below) we have inherited useful knowledge about nature which has been tested by trial and error in the course of evolution of many generations. Fatally wrong "assumptions" were extinguished by selection.

Natural science

Natural science proceeds from hypothetical realism, which postulates:

1. There is a real world.

2. Through scientific observation and experiments we can learn objective facts about it and formulate laws which generalize these observations. Although this postulate cannot be proven, it consistently works, which means that so far all observations about nature do not contradict this principle and fit together in a converging framework by complementing each other.

3. The laws of nature exist in independence of their experience (Erfahrung) by mankind, *i.e.* regardless if and when humans discovered them. In other words, science postulates that the world around us is not created and ordered by science, but the ordered world around us exists whether we perceive it or not. The conception that nature is only ordered (or even created) by research, that for instance quanta did not exist before Planck described them, leads to implausible consequences. The consistency of independent observations contradicts it.

Different people perceive different, subjective images of the world. Yet we can arrive at descriptive statements about the world that can be tested in a way that is intelligible intersubjectively.

Natural science operates by applying a number of strict game rules by which it arrives at its statements about the world. The most essential of these rules is that a phenomenon is only regarded as objective if it can be observed by several independent observers (*i.e.* intersubjectively). Natural science, too, can only observe properties and it can subsume combinations of properties in its concepts (Begriffe). Implicitly or explicitly all terms of natural science are defined by procedures of measurement (which include counting and all kinds of observable, measurable tests). But we also know of phenomena inaccessible to natural science, our feelings and emotions (Gefühle), our primary, subjective experiences.

Values in science?

The acknowledged scientific regularities or laws are objective, also in another sense: They do not depend on values, ideologies, or beliefs. The laws of the conservation of energy (Energiesätze), the equations of motion (Bewegungsgleichungen), and the laws of heredity are simply generalized objective descriptions of observations, as are Kepler's Laws, Ohm's Law *etc.* They can be verified intersubjectively and reproduced under defined conditions at any time. Scientific laws express approximations of reality; within their well-defined limits of accuracy we cannot choose to believe them or not. But this only applies to ascertained statements, not for working hypotheses, no matter how plausible they may appear. Scientific certainty is approached on an asymptote.

It has become fashionable to claim that natural science is not free of ethical values and that scientific findings are influenced by the formulation of the questions. This is a fallacy. This way, the process of gaining knowledge is confused with its result: What depends on systems of values and on fashions is the object of interest and consequently the decision of what is researched, – not the experimental results. In certain times certain questions gain a priority of interest for reasons of ideology, politics, or a need of the time (*e.g.* space research or the development of medicine or weapons), because new methods become available, or even as a kind of fad.

Value judgments and values are alien to the statements of natural science. This does not prevent value judgments from entering the process of scientific research in the formulation of questions and working hypotheses. But since science requires objective proof before it accepts statements, this can only have temporary effects. For instance, not even the massive exertion of Nazi or Soviet influence on biological research had a lasting effect on genetic science. Laws of heredity that were verified through experiments (*e.g.* Mendel's) are independent of conceptions of values and ideologies. Ideologies and conceptions of values can only hinder and delay research and acceptance.

Premature assumptions

Another thing that depends on the spirit of the age (Zeitgeist) is that some tenets (Aussagen) are already believed when there is no hard proof yet. One much-discussed example (*e.g.* by Paul Feierabend, 1983) is that at the time Galileo discovered the moons of Jupiter, telescopes were not good enough to make them out beyond doubt. That may be so. But today we have sufficient tools of observation at our disposal to indisputably observe the moons Galileo described. Throughout the history of science, many early claims have been refuted through the accretion of specific knowledge, while others were verified. Confirmation is preceded by a phase of search and conjectures. However, there is no clear borderline signalling when in a process of research scientific confirmation is reached. One has to be aware that scientists' personal convictions may be wrong, – plausible and still not true. But even false assumptions may lead to a right path to a solution of a scientific problem. On the long run only the value-free evidence counts – regardless of who found it and for what reason.

The power of science

Part of the strength of natural science stems from its limitation to what it has access to. Science has specific game rules and within them it is powerful. If a scientist oversteps this area, then it is as if a chess-player made forbidden moves. The sciences' process of determining the truth is protracted and can also reach its goal after going astray for a time. That is why during the process of decision it is not always clear whether one has arrived or not. In science's inductive method of proof a statement becomes more and more probable with the accumulation of evidence and is finally regarded as certain, because it has invariably been confirmed in many tests. The progression to a secured statement is very gradual. Unlike the deductive proofs of mathematics which proceed logically and cogently from unprovable axioms, scientific proofs are inductive, they have no sharply defined line beyond which they can be regarded as certain.

Because of the rigidly applied rules, a number of hypotheses can be refuted unequivocally. For example, Georg Ernst Stahl's (1660–1734) highly original theory which assumes that combustion is based on the release of a substance 'phlogiston' was conclusively disproved by the experiments of Antoine L. Lavoisier (1743–1794) and never again discussed thereafter. It is different in the humanities. These always retain their old theories as possible alternatives. The perspective of the moment determines which theory is deemed appropriate to the time. Unlike scientific controversies, those in the humanities frequently cannot be resolved by newly-gained information. Humanities change not so much in the growth of expert knowledge (Fachwissen), as in changes in the province of differentiation of meanings (Alexander, 1988).

Psychology is not a pure natural science. Some psychologists think that the restriction of science only to objectively observable, measurable terms is too reductionist. They use a different approach. They attempt to describe, approach, and define a phenomenon such as creativity or adaptation from a multitude of aspects; that way they hope to grasp the totality of the phenomenon's properties (Myers, 1983).

What does biology know of the nature of living beings?

Biology is the natural science of living beings. It does not deal with all aspects of living beings; it is only concerned with those properties that can be observed objectively, *i.e.* quantified, counted, mediated intersubjectively, and verified. This is also true for ethology (Verhaltensforschung). Of course one can deal with biological objects in different ways. Psychology and sociology do so with a variety of scholarly methods, some strictly scientific, some not (nicht naturwissenschaftlichen aber wissenschaftlichen Methoden). I may even be convinced that I understand the feelings of my dog; but that is not natural science.

Consciousness (Bewußtsein)

My consciousness is my primary subjective mental experience, my feelings and emotions. But I know that solely from self-observation and because of this I cannot prove whether other beings have a consciousness or not. It seems reasonable to assume that other humans have the same or similar conscious experiences, and we are supported in this assumption by the fact that we can communicate by language about them. Language is our main access to study another person's consciousness. However, we never know for sure if other people's descriptions of experience are true or lies. With regard to animals we can only guess, and our guesses become more and more speculative the more distantly we are related to them. That chimpanzees and gorillas experience sensations in some kind of consciousness appears fairly certain to us. When it comes to the housefly, the mosquito, and the slug, any guess – be it positive or negative – is pure speculation. Experiments involving their own mirror image indicate it is likely that chimpanzees experience a consciousness; however, the negative results of similar tests with other animals such as dogs cannot exclude the possibility of them having a consciousness. Epistemology does not provide a criterion as to whether or not a beetle has a consciousness. Therefore the statement that bees have no feelings is as unsure as that to the contrary. I can conceive of no scientific way to measure subjective mental experiences, consciousness, and emotions objectively. Some biologists, uncritical in their application of the scientific method, tend to disregard these limitations of natural science. In psycho-physical experiments we can fairly accurately quantify our subjective sensations as evoked by various stimuli (v. Campenhausen, 1993). By observations of behavior and reports of other people's self-observations, we are able to extrapolate general statements about human feelings, *e.g.* that consciousness is linked to attention and short term memory (Crick and Koch, 1990; 1992). We do not doubt that other people have similar feelings when they speak of identical conscious experiences. Again, this principle fails with regard to the bee.

We know only very little about the development of our consciousness in our individual life, since our memory of our early days is very poor. We feel justified to assume that consciousness develops during the early phase of our ontogeny, but we cannot positively decide when it starts.

Proceeding from the scientific finding that in the course of evolution organisms evolved from more primitive forms to higher, more differentiated ones, one may ask the question if there have been earlier stages of consciousness in the evolutionary forbears of humans. Furthermore, are there less differentiated stages of consciousness in contemporary animals? Even though natural science is unable to make statements on consciousness, the multi-disciplinary answer to both questions is probably 'yes' (Delbrück, 1986). It seems plausible to us to assume that in higher animals with their (in comparison to humans) less highly evolved brains, there possibly exist earlier stages, less differentiated forms of consciousness. Apes, for example, are able to identify themselves in a mirror, so it is probable that – unlike monkeys – they already have an individual consciousness (a self awareness).

The hypothesis of psycho-physical identity

In my subjective experience, physiological processes in my body, *i.e.* processes that can be physically measured, are closely linked to events in my consciousness. If I get burned on a hot stove, I feel pain. We know – *e.g.* from psycho-physical research during brain surgery under local anesthetics – that the activity of the human brain is accompanied by certain reported psychic experiences, sensations and emotions. Certain patterns of excitation in the brain then correspond to certain feelings. Reports by humans of their feelings and sensations when specific areas of their brain are subjected to focused stimuli, and of deficits experienced due to brain injuries allow us to identify a relation between measurable brain activities and certain – reported or even self-experienced – conscious experiences. We can relate localized brain activity that we measure physico-chemically to sensations human subjects report. Such investigations have shown that conscious experience is linked to activities of neurons in restricted areas of the cortex, the primary sensory fields, and not to activities in the many other fields. We consciously experience the shadows of reality (Creutzfeld, 1986). We can identify the activity of certain groups of neurons when test subjects report certain hallucinations. The reports of hallucinations (phosphens *e.g.*) can even be used to identify the circuitry in the brain (Ermentrout and Cowan, 1979; Cowan, 1986).

These findings lead us to the multi-disciplinary postulate: Consciousness, feeling, and thinking are functions of the brain. However, none of these important experimental observations can give natural science a direct access to consciousness and feelings. This is a border area of science: In ethology different observers observe the same behavior of a subject in a given situation; a classical scientific experiment. In psychology we can have reports by different subjects on their mental experiences as observed individually under similar experimental conditions. This latter case meets several – but not all – criteria of science. The two cases are different insofar as the primary mental experience can only be observed by the person concerned (not several independent observers) and that the reports are "translations" which may be (possibly even systematically) adulterated. Of course one can conjecture and successfully study how consciousness is linked to physical processes, but that is pure natural science no longer, rather a promising, very interesting multi-disciplinary type of scholarly research (Wissenschaft). Such an approach to acquire knowledge of consciousness by combining neural and psychological data from experiments on humans and animals with subjective reports from humans is termed "triangulation" (Churchland, 1986; Dennett, 1991; Flanagan, 1992).

Limitation of scientific brain research

In my conscious life I experience that my will is the cause of my actions. As stated above, although my primary subjective experiences, mental experiences cannot be an object of natural science since they cannot be observed by independent observers. Natural science however, proceeds from the postulate of causality. It always requires causes that can be registered with its specific, *i.e.* physical or chemical, methods. Consequently a mental process, *e.g.* will, may not be given as a scientific cause for an action (see also Bieri, 1992). Science demands uninterrupted causal chains and will cannot be an element of such a chain. It cannot substitute but only be associated with a physical cause. This has been confirmed so far in every investigated example. In cases where biology neglected this principle, progress in scientific understanding was not promoted. A prominent example is that of vitalism (Hans Driesch, 1867–1941) which assumed that the not observable "entelechy" is responsible for the ordered development of an organism. This theory led to a dead end. In the meantime the mechanisms of development have been explained scientifically and no assumption of an entelechy is needed.

Scientific brain research is thus constrained by the dictates of its method and is unable to make statements about certain feats of the brain. A philosophical no-man's land stretches between the areas in which brain researchers can reliably operate and the area of "body, mind, soul, and consciousness". Scientists scrupulous in the application of their methods are lost here, while philosophers seem to have still no better means to bridge the gap than Kant (1724–1804) had in his time. On this subject Emil Du Bois-Reymond (1818–1896) wrote in 1872: "ignoramus et ignorabimus", *i.e.* "we don't know and we will not know". Wittgenstein (1889–1951) described the inexplicable axiom that certain constellations of nervous activities gain the quality of subjective consciousness as "the mystical" ("das Mystische"), (Wittgenstein, 1961). However, Flanagan denounces those who continue to believe this as "the new mysterians" (Flanagan, 1992).

I do not know of any methodically stringent solution to the so-called Mind-Body Problem from pure science. Science sees the physical and chemical processes within an organism as logical, consistent, and unbroken causal chains in themselves. There is no room for consciousness, emotions, or will to intervene in the organism's physical machinery. On the other hand, our actual primary experience is the exact reverse, namely that our will causes our action. We move because we want to. It seems we are dealing with two different aspects of the same system, of which science can only see one.

This situation can be compared metaphorically to looking at a black-and-white photograph of an unfamiliar colored painting, *e.g.* one by Picasso. The black-and-white photo depicts the painting in its entirety, but will leave us unable to make positive statements about the colors. This metaphor, like that of the projection of a three-dimensional object onto two dimensions, is of course a poor comparison, but it may help to illustrate the problem.

There is another analogy: If one studies the hardware and the performance of a computer which is programmed to execute a complex program it may be practically impossible to find out the purpose and the function of the program. Braitenberg (1986) points out that even in relatively simple robots very few insights about their program can be deduced from observing their behavior. Thus consciousness, mind, and will may be connected to something behind the physical performance of the brain.

Misunderstandings

That the brain is the seat of consciousness is not a scientific, but an multi-disciplinary conclusion. This issue has recently led to a number of – mostly futile – discussions.

There are other misconceptions that derive from the failure to keep the methods of natural science and the humanities properly separated. In detailing the differences between the human brain and those of animals people often claim a number of deficiencies of "the" animals, many of which are not established yet – or not determinable. Even some biologists maintain that morality does not exist in the animal world (Vogel, 1988). This statement is just as inadmissible and "chauvinistic" as the statement that only humans have a consciousness. By strictly scientific methods it cannot even be proved that the behavior of humans is moral. For the judgment of morality it does not matter whether or not a behavior is genetically advantageous, *i.e.* if according to the "selfish gene" concept a parent is favoring its own offspring (Dawkins, 1976). Morality is a matter of psychic motivation, to which natural science has no access. Consequently it is impossible to prove or disprove scientifically that animals act from moral or altruistic motives.

A similar argument holds for free will. A free will decision is subjectively intentional and neither predetermined nor random. Again, there is no scientific method which allows us to decide whether a certain decision has been made on free will or not. We cannot even safely decide whether free will decisions exist at all. Nevertheless it seems most comforting to believe that we have a free will.

Current status of brain research

Over the last thirty years, the introduction of new methods allowed great advances in the research of simple and more complex brains (*e.g.* those of mollusks, insects, and crustaceans, various vertebrates and humans). Brain research is conducted on four levels. The investigation of all four levels is necessary for the understanding of the brain. Each of these levels has to be investigated with its own specific methods by which a functional understanding at that level can be reached. Each separate level represents a sort of a closed area of description (Beschreibungsraum), complete unto itself. However, the knowledge of the properties of the elements alone may not be sufficient (and does not even have to be complete) to explain the properties (and their possible changes and development) of the system. The properties of the system are determined both by the properties of its elements and their functional linkages. On the higher level only the input-output dependencies of the sub-systems are important, not the manner of their actual realizations. Similarly, the knowledge of the physics of elementary particles is not yet sufficient to understand all properties of solids (Festkörper).

Four levels of brain research

1. The molecular-biological analysis of processes in the brain has reached a stage where several key reactions can be described on the molecular level (Fischbach, 1992). For example, in the synapses, the computing connections between neurons, certain protein molecules, the transmitter receptors, are of great interest. By changing their conformation they act as switches. There are two types of such receptors: a) Ligand-gated ion channels (*e.g.* acetylcholine receptors) which change the conductance of the subsynaptic cell membrane. b) Modulators (*e.g.* dopamine receptors) which modulate the performance of the synapse via the control of amplifying (G protein-mediated) enzyme cascades. Many of these receptors are already chemically characterized in amino acid sequence and molecular shape, and subtle subtypes of functional importance have been determined.

Another interesting type of protein molecule is that of associative or convergence-detecting molecules (*e.g.* certain N-methyl-D-aspartate (NMDA) receptors and adenylate cyclase) which play a key role in memory formation. They have the property to detect coincidence of two independent biochemical signals originating from the activity of different neurons. If two converging neurons are active in a specific time sequence, such molecules are activated more strongly by the combined action of the two pathways if these actions coincide. One example of such a protein's role is the classical conditioning learning in the gill-withdrawal reflex of the snail *Aplysia*: Here the excitatory effectivity of a sensory neuron's synaptic ending is influenced by the presynaptic contact of a modulatory neuron. When the level of intracellular calcium in the ending of this sensory neuron is raised (due to the conditioned stimulus) shortly before serotonin is released by the modulatory neuron (due to the unconditioned stimulus), the activation of the convergence-detecting enzyme adenylate cyclase is strongly enhanced. This causes an augmentation of the transmitter release of the sensory neuron that can last for hours (Kandel and Hawkins, 1992).

2. Cellular level. Processing information means the performing of mathematical operations, algorithmic transforming, and combining of sets of data. This task is performed by the neurons in the brain. *Neurons* are interconnected, sophisticated processing elements. Their performance, *i.e.* their admittance and amplification factor is variable:

a) It can be modified (adjusted) according to use (experience), *i.e.* it can adapt.

b) It depends on the state of activity of neighboring elements, taking into account neighboring activity patterns (cooperativity of local nerve nets).

c) It can be modified by other systems of the brain which regulate sleep, alertness, and attention or represent emotional states.

Synapses are the calculating elements of the neurons; the single synapse operates linearly to a large extent. The firing, *i.e.* the generation of nerve impulses, of a neuron is determined by the voltage across its postsynaptic membrane. When a neuron is depolarized, *i.e.* this voltage is lowered below a threshold value, one or several spikes are fired. The many synapses of the neuron can modify this voltage. The kind of synaptic influence (positive or negative) and its degree (Wertigkeit) depends on the morphological and the biochemical characteristics of the synapse concerned. The larger the synaptic contact zone, the greater is the influence of that synapse. In addition, the performance of synapses can be modified by neurohormones (chemical modulators). Excitatory synapses depolarize, while inhibitory synapses hyperpolarize or stabilize the membrane voltage. Using their specific combinations of synapses neurons are able to add, subtract, multiply, divide, differentiate, integrate, detect coincidences, etc., and to perform all kinds of logical operations (Nicholls *et al.*, 1992; Shepherd, 1993).

3. Connected groups of neurons. At this level, the investigation of specific brain regions has resulted in the detailed description of functional linkages between groups of neurons, *e.g.* reflex arcs, feed-back loops *etc.* The architecture of the mammalian cortex turned out to be surprisingly uniform. The building units are module-like local nerve nets which all process information in principally the same way. They form a continual series of parallel vertical loops, *e.g.* senso-motor control loops. Horizontal connections between these loops at various levels make possible integrated responses to a stimulus *i.e.* adequate behavior. Among other things this has been most successful in tracking the neuronal connections starting from the eye to many regions in the brain of vertebrates and insects. For instance in primates the visual tract is divided into several separate pathways from its very start in the retina. Processes of neurons with similar functions are bundled and pass through the same relay stations. Thus motion-sensitive neurons, color-sensitive ones, and two groups of different shape-sensitive tracts are all grouped separately. Although interconnected, they are routed separately and switched-over at five to six relay stations, four of them in the cortex. The various brain regions concerned are already characterized to a degree in respect to their functions. *E.g.* there are groups of neurons in the area V4 of the visual cortex which preferentially respond to stimuli which are oriented on the retina. Each neuron has its preferred stimulus orientation and close neighbors have a more similar orientational preference than more distant neurons. In other areas of the cortex, *e.g.* the primary visual cortex (area V1), we find topographical maps which unambiguously correlate the locations of sensory – or effector – cells in the periphery and the locations in the cortex to which the tracts from these neurons lead. Adjacent sensory cells here project to adjacent neurons in the cortex. So in area V1 of the visual cortex we find a distorted map of the array of the visual cells of the retina (Livingstone and Hubel, 1988; Zeki, 1992; 1993).

Another example is Hebb's concept of cell assemblies (Hebb, 1949); according to which neural brain function is based on patterns of coordinated activities of many cooperating neurons. In such a cell assembly excitation can circulate for a while and form a reverberatory circuit, thus providing reinforcement as a prerequisite for learning. Repeated coincident activity of two converging neurons strengthens the concerned synapse.

3a. Language is a property typical of humans. As a rule, human language is regarded to be associated with consciousness. The question of the contribution of consciousness for language poses itself as in the other forms of behavior. Has consciousness a special role in language? I do not see a purely scientific way to decide. However, there are many aspects of language which are accessible to objective investigation by independent observers. Language is an instrument used to communicate concepts, experience, and other information among individuals. Human language is primarily sound-based, but there are forms of language which use other types of symbols, *e.g.* sign languages use viso-motor signs (or letters in writing). There are more primitive forms of language used by communicating social animals (*e.g.* bees) but in humans it is a powerful tool of a sophistication unreached in the animal kingdom, which can express and transmit abstract concepts. The neural base of human language can be described by three functional systems (Damasio *et al.*, 1990; Damasio and Damasio, 1992):

a) Concept representation: The brain has the property to categorize passive and active experiences, *i.e.* sensory input and motor output patterns of activity, according to brain-specific categories such as shape, color, intensity of action, sequence, *etc.* It subsumes observed properties which it regards as connected to syndromes ("features"), phenomena on a higher level of representation. In this way it subsumes different experiences to recognize objects, events, and relationships. Furthermore it extracts homologies and relationships between objects (features), *i.e.* between received patterns of sensations (experiences). The brain of many mammals can recognize and use symbols. The human brain can model reality in symbols. This competence to recognize homologies and to use symbols is prerequisite for the use of language.

b) Word formation: Furthermore, the human brain possesses a set of structures which provide the special tools for language. Generally located in the left cerebral hemisphere there are centers to represent lingual sounds (phonems) and words (morphemes) and (syntactic) rules how to combine them to sentences. These serve both for generation (production) and reception of speech. Words and rules have to be learned, but the basic structure of the human language (*e.g.* the use of symbols and the distinction of nouns on the one hand, and verbs or functors on the other) seems to be provided innate in the brain.

c) Mediation: As for other actions, there are control systems in the brain which organize the proper combination of concepts and their expression in symbols, words, and sentences.

These three systems can already be roughly localized in specific brain structures where lesions produce characteristic deficits, *e.g.* in word finding, word understanding, and grammatical processing.

4. In the domain of systems analysis the neurophysiological base of the behavior and learning achievements of entire organisms become visible. So far, this has most successfully been done with a number of species of invertebrates (Reichardt, 1982; 1986; 1987) but becomes increasingly successful also in vertebrates, *e.g.* amphibians, birds, and mammals (cats, rats, and monkeys). One example is motion detection and flight control in flies. The "wiring diagram" of the neurons from the retina of the fly has been mapped in detail (Strausfeld and Nässel, 1981); its connections within the brain and much of its functional significance is now understood (Hausen, 1993). The search for the algorithm for detection of motion started from a detailed behavioral analysis of the optomotor response of insects. This led to a correlation model and further step by step, alternating with experimental verifications, to a quantitative description of a control system capable of accounting for the main features of fixation, tracking, and chasing in flies. The equations can predict in a satisfactory way the trajectory of one fly chasing another (Hassenstein and Reichardt, 1956; Poggio and Reichardt, 1976; Reichardt and Poggio, 1976; Egelhaaf *et al.*, 1988; Egelhaaf and Borst, 1993).

Uses of brain/computer analogies

Brains and computers differ fundamentally in their structural components. Computers are made of electronic elements, whereas the brain is of "flesh and blood". Therefore a direct structural comparison of brains and computers is not very profitable. However, the functional principles of both systems can be compared and exploited for brain research and computer development. After having characterized a brain function experimentally it can be tested whether the resulting principle works in computer simulation which may even lead to the construction of robots (Franceschini *et al.*, 1992; Martin and Franceschini, 1994). On the other hand, conceptual models and computer development have contributed significantly to advance our understanding of the brain. Information theory and cybernetics have supplied conceptual tools. An example is the figure-ground discrimination. Flies can detect spatial discontinuities of speed within the retinal images and thereby distinguish figure from ground on the basis of relative motion. The interplay of experiments and theory led to a model for the algorithm used by the fly's visual system, which discriminates a figure from its visual context by laterally inhibiting nonlinear interactions (of order 4) between motion detectors (Reichardt *et al.*,1983; Reichardt, 1986). This principle, discovered in insects, is now established as a general principle for the processing of visual information in a variety of species including man.

Important functional principles have been discovered through comparisons with computers. The "game" of exploring computer capabilities can help to discover functional principles which later may be detected in the brain. Strategies for learning in networks of modeled neurons have been explored in computer simulation (Hinton, 1989; 1992). The competitive learning algorithm in the modifications of T. Kohonen and H.B. Barlow is able to simulate and explain certain characteristic properties of the brain as feature categorization, formation of clusters of neurons with related functions, and determination of decisions by averaging the outputs of relatively small numbers of neurons. The principle of this learning algorithm is based on a large number of inner model neurons (so-called hidden units) competing in order to represent best the input pattern. By learning (self-adjustment of its "synaptic weights"), the winner unit modifies itself to improve the representation of the particular input pattern. If the competing hidden units are uncorrelated, this learning is not restricted to a single winner unit and each input pattern is represented by a cluster, a small group of selected hidden units.

It is, however, important to pay attention to the usefulness of model simulations. A model must be disprovable by critical experiments; a model which can describe too many possible outcomes of test experiments is of not much use. Also an analogue, e.g. a computer simulation which merely provides the same input/output relation as a certain brain function, is only of limited use, because it may not tell us how the brain does it. It can of course be used if one only wants to investigate how the simulated functions are handled as modules at higher levels of integration. If, however, a simulation contains relevant elements and representations of units and processes which are actually proven to be present in the particular brain process, it may indeed help us to understand how the brain works.

How the brain processes information

Information is processed and transmitted by the neurons of the nervous system using their adaptable connections. Information can be expressed in analog or digital encodings. First it was thought that the brain encodes information digitally because the nerve impulse, the spike, superficially resembles pulses used in digital computers. The idea was that the brain uses two different discrete characters, spike and no spike. However, it soon became clear that the information encodings of the brain are better described as a kind of analog encodings. Consistently identical nerve impulses serve to mark the variable length of intervals between them; the intervals contain the information, in other words, the information is encoded by pulse frequency modulation.

For a long time it was a mystery why the brain can process some massive information so fast even though the individual steps of its operation are relatively slow, taking about a millisecond each. More detailed analysis has revealed that the brain processes much of the information in parallel fashion. For example, a visual image that is received by an array of about 100 million visual cells within an eye, is encoded and then routed to the brain by means of the optical nerve, which consists of about one million parallel pathways for each eye. This massively parallel encoded visual information is transmitted to more than fourteen different locations within the brain, where it is processed simultaneously – *i.e.* also in parallel. In this process, the visual information is decoded and distributed according to brain-specific categories, manifold represented at several different locations in the brain. For instance, there are certain groups of nerve cells in the brain that only respond if the light stimulus belongs to a specific color group. Another type of neuron ascertains whether an optical stimulus is a line of a certain orientation in the visual field, yet other types of neurons respond only to moving stimuli, and so forth. Through its own categories the brain collates stimuli of similar significance to the organism. Thus, colors are brain-internally constructed categorizations (and subjective sensations) based upon objective properties of our environment. They are the results of our brain's computations, physical stimuli sorted out according to its peculiar brain-specific categories. Physically very different stimuli can cause in us virtually the same perception of color (Farbempfindungen).

After this parallel evaluation, the information from various brain locations is combined "expediently", *i.e.* according to the needs of the organism. This results in patterns of excitation which finally control the behavior. So organisms react to what may be relevant in a given situation. The so-called binding problem, namely how the brain manages to assemble and combine correlated information concerning the different aspects of an object, and how information concerning other objects which the brain processes at the same time is kept separate, is still not well understood. Synchronous or oscillatory activities in groups of neurons may serve as labels to recognize evaluations which belong together (Freeman, 1988). Abeles proposes 'Syn-Fire-Chains' of neurons which can reproduce spike patterns with very high accuracy even after long delays. In different neurons of monkeys engaging in learned complex behavior, he and his coworkers observed coherent, exactly (± 1 ms) reproduced neuronal activity patterns with different time delays of up to 500 ms. These specific spatiotemporal activity patterns may be used to recognize the related activities of multiple syn-fire-chains in different cortical locations in order to combine them (Abeles *et al.*, 1993). Singer and his coworkers describe 40 Hz oscillations, which are observed to occur simultaneously in several separate areas of the brain while processing different aspects of a stimulation of the eye of a cat (Engel *et al.*, 1993; Singer, 1994). Freeman suggests an alternative explanation: "Groups of neurons display chaotic behavior, that is, their firing pattern seems random but actually it contains a hidden order. Like all chaotic systems, these neural patterns are extremely sensitive to minute influences. The sight of a familiar face, therefore, can trigger an abrupt shift in the firing pattern corresponding to a shift in one's awareness" (Freeman, 1994).

Our brain is a parallel processor of information encoded by pulse frequency modulation. The algorithm which is used by the working brain when solving problems seems to be different, however, from the one which we consciously experience when we introspectively analyze our human thinking (Creutzfeld, 1986).

The role of chance (Zufall) in brain and computer operations

The combination of chance and selection is the basis of evolution (Monod, 1971; Mayr, 1979; Eigen, 1987; Stieve, 1990). It enables systems to become fit to many unforeseen challenges. It may be possible that the brain uses randomness not only in development but also for certain decisions, *e.g.* random choices in searching behavior or trial-and-error learning (Gierer, 1985). In computer application chance is often used (*e.g.* in so-called Monte Carlo and artificial life simulations). In connectionistic computers chance is additionally used to adapt the performance to the desired tasks. However, up to now in computers real randomness is usually not achieved. The so-called random number generators in reality produce deterministic numbers, *e.g.* the digits proceeding from the calculation of π (pi). Stochastic simulation on computers is used as a tool for the performance evaluation of complex communication networks. Yet in certain cases it has been shown that employing these numbers leads to wrong results. The use of real random numbers, instead, *e.g.* from radioactive decomposition or from electronic white noise have led to very good results (Richter, 1993). It does not seem impossible that the brain uses real randomness by amplifying atomic processes which are determined by quantum uncertainty (Heisenberg's uncertainty principle; Gierer, 1985).

Learning

Patterns of behavior of an individual may be preprogrammed, *i.e.* innate, or they may be learned in specific situations. The structural and molecular base of the neural memory responsible for the capability to learn are as yet barely understood, – but in contrast to subjective psychic experiences, they can be researched in an objective scientific manner.

A system is able to learn if it can improve its performance through experience. Memory is a prerequisite of learning. Biology distinguishes between two kinds of learning processes:

1. Individual learning: An organism utilizes its individual experience for more adaptive behavior. [What is called "learning" in Artificial Intelligence (AI) research is a simulation of individual learning]. The capability for individual learning is hereditary.

2. Superindividual learning: Individuals more suited to the ambient environmental conditions can reproduce and pass properties on to the next generations with greater success, *i.e.* this selection-based type of learning by trial and error (which is also found in the immune system) is achieved not by single individuals but by successions of generations.

A combination of these two learning processes gave rise to the evolution of fairly competent organisms; organisms which are not optimal but fitter than their local competitors (Evolution often leads to a "local optimum" which is not unlike a metastable state in chemistry; Eigen, 1987). There are animals which have a large part of their inventory of capabilities fixed in the programming of their brains, "hard-wired", as it were. But often it seems to be more advantageous for organisms to be able to learn in order to adapt quickly to unpredicted new situations. Houseflies seem to learn very little. Bees learn in a specific, narrowly limited field; feeding, building honeycombs, communicating with fellow bees, and a complicated mating behavior are pre-programmed, but where to find a productive flower can be learned. Mammals, including humans, must employ their learning capability to a large extent.

The classic test to decide whether a behavior or performance is an innate, instinctive activity or learned from others, is the so-called Kaspar Hauser test. For this an animal has to be raised in isolation and all relevant influence of others excluded. For ethical reasons, this of course cannot be done with humans. Therefore the question of instinctive activities, *i.e.* which part in human behavior is innate and which is learned, can only be approached indirectly and the answers are less secure.

Learning, *i.e.* adaptive changes in performance due to environmental conditions, starts during early infancy or even before birth. The development of brain architecture greatly depends on experience and exercise. The visual experience during the infancy of an animal influences the connections of the concerned neurons in the brain (Wiesel, 1982; Singer, 1990; Nicholls *et al.,* 1992). But due to the great plasticity of the brain the nervous connections are continuously modified even in adults. Use and non-use affect the structure of the brain; hereby use leads to improvement, while non-use causes deterioration, even withering of connections. Neurons seem to compete with each other for connections, and thus influence. The implications of this "Neural Darwinism" are discussed by Edelman (Edelman, 1992). It was shown in adult monkeys that the areas which represent the fingers in the topographic maps of the somatosensory cortex depend, and change continually, according to use and non-use (Merzenich, 1989). Possibly, one of biology's greatest unsolved riddles, the structure and mechanism of our memory, its storage and retrieval, may be solved in not too distant a future.

Parallel and connectionistic computers

Whereas brains have developed in the course of evolution without clearly defined requirements and parameters for their performance and were tested by a near-infinite number of trials, computers are designed for certain purposes and optimized under a tight budget and schedule. On the other hand computer development has been greatly stimulated by brain research; it used several concepts of brain functions (such as Hebb's learning rules, parallel processing of information, and neural networks).

Modern, high performance computers use discrete digital characters to encode information. At first, computers were developed on the principle of sequential data processing. Sequential computers – unlike our brain – compute information one step after another. This step-by-step processing is governed by a system clock. In the course of technical development the individual steps became faster and faster, much faster than in the brain. A single nerve impulse lasts for nearly a millisecond; consequently the highest frequency at which the brain can transmit information is less than one kHz. Modern high performance computers operate today at a clock rate of 100 MHz or higher, *i.e.* more than 100,000 times faster than the brain. Although the highest frequency used to transmit information in the brain is so low, the brain can still detect time differences of 10 microseconds by exploiting delay time differences (Laufzeitunterschiede) in neuronal pathways of different length and neurons functioning as coincidence detectors (Konishi *et al.,* 1988; Konishi, 1993; Wagner, 1994).

Within the last ten years, technology is employing the principle of parallel computers in a fertile line of development. These computers work by the principle of parallel distributed processing, *i.e.* they divide a mathematical problem into component tasks to be processed individually in parallel; subsequently the results are put together (Churchland and Sejnowski, 1992; Männer and Lange, 1994). Such computers have advantages over sequential ones. Their development has immensely changed our conception of computers' abilities. Presently using as many as a thousand interconnected processors they have achieved dramatic increases in computing power, although they still do not approach that of the human brain (Kurzweil, 1990).

Computers can be classified into four classes according to the structure of their hardware:

1. The classical sequentially working "single processor computer" ("von-Neumann computer"). The personal computer in principle is a smaller and slower version of this type (a few personal computers have two, not just the usual norm of one processor).

2. The computer with several processors. To date this type has been developed as far as containing up to 64 very fast and sophisticated processors. The most capable (and most expensive) models of this type are often called "supercomputers" (*e.g.* CDC, Cray, IBM). They are primarily suited for "number crunching jobs", especially if problems can be described in form of matrices which can be divided into many parallel calculations fairly easily. A compiler program distributes the subordinate tasks among the processors.

3. The massively parallel processing computer (MPP) has a large number (normally several thousands) of relatively simple processors which are connected amongst each other by so-called interconnection networks in order to solve a problem cooperatively in a multiplicity of parallel blocks. This type is much less expensive for its high processing capability. It includes "single instruction, multiple data" (SIMD) computers which use a single program on parallelly structured data and "multiple instruction, multiple data" (MIMD) computers. In the latter case several parallel sets of data are processed simultaneously by different programs. Today a main focus of research is to optimize the way the operating system automatically divides a complex task into parallel processor steps and then combines the partial results.

A somewhat similar (but less capable) structure can be achieved by interconnecting a large number of personal computers or workstations by means of local area networks (LANs). The main difference is that in such loosely cooperating systems each unit has its own operating system and the cooperation is effected by an additional network operating system (NOS).

4. The hardware connectionistic computer. This group of computers is based on the development of connectionistic processors which often are designed in imitation of some properties of neuronal networks of the brain. Here these neural network properties are realized in the hardware. Such "neural network chips" have hundreds or thousands of single-processor functions ("knots") per chip and many chips can be linked to form very complex networks. Properties like learning capabilities are realized in the hardware in that the connections concerned are physically modified.

(Connectionistic computer systems are mistakenly called "neural networks", even though they are much less complex than brains and as yet only utilize very simplified models of some properties of nervous systems. The term "neural network" anticipates something which in most cases has yet to be proven. But we probably will have to live with this "trade-mark" name as it already has gained general currency.)

In sequential computers, programs can also be realized which simulate the parallel distributed processing of the nervous system, such as pattern recognition or learning *etc.* Though the performance is much slower than in comparable parallel computers, the experimental conditions are easier to vary.

Computers, especially the connectionistic kind, can have a learning capability in that calculation processes leave traces in the circuits concerned which accumulate through repetition. That way they can adapt their functions to recurrent tasks. This adaptation can occur deterministically or statistically in non predictable ways. The ability of a computer to learn can either be implemented in hardware, *i.e. via* the modification of physical circuit properties (*e.g.* the resistance of a particular interconnection), or it can be achieved through the appropriate modification of software (*i.e.* program) variables. (It is also possible to achieve technical learning by means of software on a conventional, sequential processing computer, but with a considerably slower performance.)

Connectionistic computers resemble the brain also in other aspects: Memorized information in such a network of computer elements is spatially distributed. In case of local destruction, *i.e.* if part of the elements is put out of action, the total memory is nevertheless preserved – if less precise in detail. Moreover, connectionistic computers can use probabilistic elements, similar to the data processing synaptic elements in the brain, whereas digital computers use deterministic elements.

A difference between the brains and today's large computers lies in the number of input and output channels and their connections. The human brain has roughly up to 2 million sensory inputs and about 100,000 outputs, numbers which are much larger than any of today's computers which normally do not exceed 100. However, computers can compensate for this by rapidly scanning many channels (time-multiplex), *e.g.* by which the CM5 of Thinking Machines can effectively control more than a million parallel input and output channels of 64 kbit/sec (Gelernter, 1992).

During the last twenty years, computer developers increasingly borrowed from certain properties of nervous systems, which substantially advanced computer technology and led to unexpectedly successful simulations of brain properties. Progress in computer development can be expected from applying new concepts, new algorithms, *e.g.* for pattern recognition, some of which can be adopted from those found in the brain. Moreover, learning could be exploited much more. A human brain needs several years of intensive individual learning in order to reach a high performance. No computer has been trained yet by a comparable amount of experience. But it is still an open question how much we can learn from the brain in order to build computers. Perhaps as much as we learned from birds' flight for the construction of airplanes?

Memory

Computers utilize information from two basic sources:

1. Programs preset by a programmer, which computers implement "obstinately".

2. Memories of computer activities; the traces left by processing their assignments. In suitable, "learning" computers the latter is used to improve their functions.

The memory of computers is based on magnetizing, electrically charging, or optically modifying some of its component elements. The process of storage can be very rapid (magnetically and optically 10 to 100 microseconds; electrically 10 to 100 nanoseconds). Most computers store information in their memory files under specific storage addresses. If a certain part of the memory unit is destroyed locally, the information stored in that location is lost. The connectionistic computers work differently, perhaps more similar to the brain. Typically they do not use digital-storage memory files. The memory lies in the reproduction of patterns of activation. There are units or groups of units which can trigger the activation of formerly adapted output patterns. This adaptation is done by adjusting their outputs (see above).

We still know comparatively little about the biophysics and biochemistry of information storage in the brain. Here too are no memory files which store digital data under storage addresses. Apparently the traces of memory are stored locally where the concerned functions are processed. Since the cooperating cells are often (seemingly diffusely) distributed over larger areas of the brain (and since *e.g.* a few 100 neurons may be used to determine a certain feature), it seems plausible that parts of this memory often cannot be deleted selectively by narrowly localized injuries. Programs are realized in suitable specific connections between neurons. Where these neurons are arranged in order, *e.g.* topographical maps, we find specific locations for certain program information in the brain. For instance, optical and acoustical space are projected one over the other into our cerebral cortex. The loss of certain parts of the brain, *e.g.* because of apoplexy, causes the selective loss of the local "hard-ware" functions. The programs are determined in part genetically, in part by experience – *i.e.* learning.

One can tentatively distinguish different forms of memory with different underlying physiological mechanisms.

a) Short-term memory, subdivided into iconic memory, which lasts for a fraction of a second, and working memory that lasts for a few seconds. The working memory is used for the short-term processing of information in evaluations and actions; in primates its operations are carried out in the prefrontal lobes of the cortex (Goldman-Rakic, 1992).

b) Long-term memory which may last for years. In humans so-called declarative or explicit learning may require only a single experience whereas nondeclarative or implicit learning usually is slow and accumulates through repeated experience. In mammals, a region at the rim of the temporal cortex, the hippocampus, is essential for learning. It apparently serves as a temporary depository (for a period of weeks to months) during the implementation into long-term memory. Destruction of the hippocampus abolishes short-term memory and prevents the storage of new information without interfering with the use of what was previously learned. Nondeclarative learning works even without the hippocampus (Kandel and Hawkins, 1992).

The processes of information storage in the brain are much slower than in computers. For short-term storage our neurons need several milliseconds; input into our long-term memory can take minutes or hours. A whole range of different mechanisms is involved: changes of calcium ion concentration in the neurons, conformational changes *e.g.* of convergence-detecting molecules, chemical modifications such as the phosphorylation of protein molecules, also the synthesis of specific protein molecules, and structural changes such as new conducting nervous connections being created, and old ones withering (Nicholls *et al.*, 1992; Shepherd, 1993; Thompson,1993). This enumeration is not meant to cloud over the fact that we still do not understand the essentials of memory in the brain.

How and how far does our brain understand the world?

Our brain functions tell us something about the physical world. As stated above, it is amazing that our brain is able to understand nature, can understand the world. Is there a good reason for this capability of the brain? Konrad Lorenz (Lorenz, 1941, 1943) and later Gerhard Vollmer (Vollmer, 1975; see also Quine, 1968) formulated the "Evolutionary or Naturalistic Epistemology", the theory of the evolution of cognition (Evolutionäre Erkenntnistheorie), a very plausible but not provable theory as to how our brain became world-understanding:

The structure and functions of a living being tell us something about the environment to which it is adapted. If one studies the anatomy and the bodily functions of a tree-living monkey, one comes to know also some properties of the forest (*e.g.* the size of the hand tells something about the thickness of the tree branches which the monkey grasps, *etc.*). The living being is a sort of a complement of certain aspects of the environment. In a similar way our brain tells us something about the physical world. It should be possible in the future to reconstruct the stimulus situation, and by that, aspects of the physical world from the patterns of activity in the brain. Brains have models of the world and use symbols to make "Gedankenexperimente". The theory of the evolution of cognition proposes that a brain which makes better predictions regarding relevant properties of the physical world in which the organism lives provides its owner with a better chance of survival than its competitors. A better understanding of how a rock falls may be of great advantage. The brain with the better theory of the physical world, or more specifically, of relevant properties of the possible environments of an organism, has a better chance to survive natural selection. By this mechanism we have inherited useful theories of nature which have been tested through many generations. The models of the world of brains of higher animals include the subject in its environment. This should be a prerequisite for conscious self-awareness (which again cannot be observed objectively).

Through the sensory organs, the brain receives information about the physical world outside and inside of the organism. There is still no good measure of the amount of relevant information. It is not sufficient to describe the amount of information in bit (Shannon's measure of the probability); one also has to consider the meaning (semantics) and the usefulness (pragmatics) of the information. It is important for the brain to handle the information it processes and stores suitably according to its meaning and usefulness.

To interpret (to "understand") our environment the brain often uses only a few relevant clues that are characteristic – and rarely mistakable – for a certain relevant situation: a sort of a recognition code (*i.e.* so-called "sign stimuli", Schlüsselreize). These clues are used to initiate appropriate reactions by the organism. For this interpretation the brain needs and possesses certain expectations (pre-knowledge) about its environment. These expectations concern regularities of the physical world, probabilistic predictions about the environment including the probable behavior of other living beings *etc.* Part of this knowledge is learned by the individual brain during the life of the organism, but much of the understanding of the environment is innate, gained by brains in the course of the evolution by superindividual learning (Lorenz, 1978).

However, there are limits to the capability of our brain. There are properties of the physical world which we can understand but cannot possibly envisage (die wir uns nicht anschaulich vorstellen können). The most popular example is perhaps the both corpuscular and wave-like properties of light. The generally accepted explanation for these brain deficiencies is that the knowledge of these properties had so far not been relevant for our evolution for which the brain was made fit. Accordingly there may be things we cannot understand and problems we cannot solve.

Another limitation of our brain is that we appear to have no realistic conception of chance; so we almost obsessively search for a direct reason of an experienced mischief. But this is perhaps a good strategy from an utilitarian point of view. Maybe such a reductionist approach is a success strategy in most cases and therefore a "natural" brain concept which is advantageous in selection. The optimistic epitaph on the tombstone of David Hilbert (1862–1943) "Wir müssen wissen. Wir werden wissen.", *i.e.* "We have to know. We will know", expresses such an approach.

Performance comparison of brains and computers

The mammalian brain has many billions (*ca.* 10^{10} to 10^{12}) of neurons, and a neuron may be directly connected to up to 10,000 other neurons. Raymond Kurzweil points out, for example, that this "biological computer" routinely performs visual image processing tasks for which a digital machine would require a processing capability of about 10^{14} operations per second (Kurzweil, 1990). The most powerful parallel processing computer available today (CM5 from Thinking Machines) achieves about 10^{11} operations per second – still off by a factor of one thousand. But this may be mainly a quantitative difference, and we do not know how far the development of computers will lead in the future.

Are there functional properties of the brain of such quality that they cannot be achieved by a computer and, conversely, are there computer tasks which cannot be performed by the brain – or at least not as well (Mainzer, 1994)? Let us regard this step by step.

What is intelligence – or perhaps more pertinently, how does one recognize intelligence? The competence to extract in a given situation the relevant from a mass of information, to combine them and draw logically correct conclusions. Artificial intelligence research strives to have those tasks done by computers. What we regard as intelligence depends very much on the regarded task. To date, artificial intelligence is in most examples still significantly lower than that of humans. However, in certain defined situations, *e.g.* in extracting relevant information from large masses of certain data, artificial intelligence can already be better than the human brain.

Alan Turing devised a test to determine whether or not a machine is intelligent (Turing, 1950). In this test persons communicate indirectly, *e.g.* by type-written questions and answers, with a system hidden in another room, – either a computer which is made to imitate human performance, or a human. The test persons then have to decide whether they were communicating with a human or a computer. Should they not be able to decide correctly, then, Turing suggests, we should attribute intelligence to the computer. Turing claims, there is no way to find out in such a situation whether the task is performed by a person or by a sufficiently programmed computer.

Are there any problems that can be solved by the human brain, but not by machine intelligence? At the current state of the art, the assumption that there are none is becoming increasingly likely.

Gödel has shown that any moderately complex system of axioms yields statements that may be self-evidently true but cannot be proved through those axioms (Gödel, 1931; Hofstadter, 1979; Gierer, 1985). Different consequences have been drawn from that theorem: Some claim that it implies that human thought can solve problems which cannot be solved by any computer. According to Penrose, Gödel's theorem implies that no deterministic, rule-based system – that is, neither classical physics, computer science nor neuroscience – can account for the mind's creative powers and ability to ascertain truth. He thinks that the mind must exploit non-deterministic effects (Penrose, 1994). Gierer contradicts this statement claiming that in no single exactly described case it was possible to prove that there is a performance of the human mind that a computer could not do (Gierer, 1985).

Turing in 1936 theoretically investigated the capacity (Leistungsfähigkeit) of a hypothetical computing device which masters all the operations of logic and of number theory, the "Turing machine". Any problem that can be solved by a finite succession of logical operations (*i.e.* algorithmically) can be solved by a Turing machine. A problem which cannot be solved by a Turing machine, cannot be solved by any kind of machine. Turing was able to show conceptually that there are problems which cannot be solved by a Turing machine (Turing, 1936).

The Church-Turing thesis is a hypothesis about the processes which human brains use. It exists in many versions and states that if a problem is human-computable it is machine-computable, or, if a problem can be solved by human reasoning (*i.e.* only by strictly applying logic and algorithms), then machines can ultimately be constructed to do it. This thesis has not yet been disproved.

Are there performances by brains that can be observed by scientific methods, but cannot be achieved by suitable computers? The brain may solve some problems not by solely applying deterministic instructions (algorithms) but by other methods, *e.g.* the use of lucky intuition. Are intuition, creativity, and flashes of inspiration achievements specific to living brains? If we try to describe intuition, flashes of inspiration, and creativity in a scientifically conceivable way, we arrive at a different conclusion. Scientific terms always require an observable, measurable test. By what measurable features can we recognize creativity?

Creativity can be described as inventing or designing something new, based on guessing and playfully trying of unknown relations. And what is a flash of inspiration (Einfall)? The unexpected, "intuitive" presentiment of relations that have not yet been explained. Intuition could thus be described as a preference for a certain decision or procedure which cannot yet be justified by logical reasoning. It may be based on pre-knowledge and on previous experience and on (brain-) system-specific preferences. Described in this way, creativity, inspiration, and intuition can of course be simulated in computers. They might *e.g.* be simulated as random tests of runaway assumptions (derived in part from "wild" hypotheses) that are triggered by associations. These random associations, however, must be restrained i.e. afterwards checked by a not too strict, but well-dosed control based on experience and logic. (By the way, creativity and inspirations need not to be right.)

The question of consciousness and subjective experiences also remains for computers (Penrose, 1989). As stated above, natural science is unable to make statements about them since they cannot be observed by the methods of natural science. Self-observation is limited to a single human individual. In psychology, one person's subjective findings are generalized and applied solely to other humans. They can be extrapolated neither for beetles nor for computers. That is to say that no one can prove or disprove that computers can have a consciousness. Maybe consciousness has some relation to information. The brain uses symbols to model the reality and manipulates them in "Gedankenexperimenten". These symbols represent information. It is conceivable that consciousness is a – not objectively measurable – property of certain complex systems, a property that does not depend on the material representing that complexity, just as information is inherent in temporal and spatial patterns and is not bound to a specific form of matter or energy. Some people, inspired by chaos research, suggest that consciousness may be an emergent – that is unpredictable and irreducible – property of the brain's complex behavior (Crick, 1994; Rasmussen, 1994). However, if consciousness is an emergent property of the brain, it is still not a property that can be observed by an independent observer.

Some postulate that consciousness (self awareness) is a quality peculiar only to the living brain. Can arguments be found for this hypothesis? Two cases are conceivable:

1. Consciousness is an additional property characteristic of certain brain functions. It is peculiar to the brain of humans and their relatives.

2. Consciousness is a property of certain data processing systems, regardless whether they are living beings or machines like complex computers.

I do not see how this alternative can be tested empirically.

Why is the performance of certain brain functions accompanied by subjective experience? Does consciousness have a biological advantage for the observable performance of an organism, that is, a bonus in evolutionary fitness (Vollmer, 1980; Barlow, 1980)? Consciousness could make satisfaction possible which may serve as a bonus or reward to approach optimal performance. However, if there were a robot which had all the objectively observable third person properties of a human, could one distinguish it from someone who has consciousness? Searle has discussed a related question in the "Chinese Room" Gedankenexperiment. He shows that one can give the right answers to questions without understanding the foreign language in which they are asked, just by applying proper algorithms (Searle, 1984). Natural science alone cannot supply an answer to these questions (Chalmers, 1994; Horgan, 1994). Some take it for granted that consciousness is not an additional property but by itself causes an improvement of measurable performance. The problem is that so far no one has come up with a method to prove or even to test this hypothesis.

Furthermore, the question if there are capabilities of the brain inaccessible to natural science that a computer cannot furnish, is in all probability not solvable.

Acknowledgments

First of all I wish to thank Jan Dieter Spalink for his advice on computers and many fruitful discussions. Thanks also go to Marlies Dorlöchter, Carmelita Görg, Karl Georg Götz, Klaus Hausen, Hans Dieter Lüke, Christian M. Müller, Walter Oberschelp, Jürgen Schnakenberg, Gerhard Vollmer, and Henrik Walter for their advice. Peter H. Marsden and Tilman Stieve helped me a great deal in preparing this manuscript in English readable both for Germans and Anglo-Saxons.

Abeles M., Prut Y., Bergman H., Vaadia E. and Aertsen A. (1993), Integration, synchronicity and periodicity. In: Brain Theory: Spatio-Temporal Aspects of Brain Functions (A. Aertsen, ed.), Elsevier, Amsterdam, New York, London, Tokyo, pp. 149–181.
Alexander R. D. (1988), Über die Interessen der Menschen und die Evolution von Lebensläufen. In: Die Herausforderung der Evolutionsbiologie, (H. Meier, ed.), Piper Verlag, München, pp. 129–171.
Barlow H. B. (1980), Nature's joke: A conjecture on the biological role of consciousness. In: Consciousness and the Physical World (B. D. Josephson and V. S. Ramachandra, eds.), London, pp. 81–94.
Bieri P. (1992), Was macht Bewußtsein zum Rätsel? Spektrum der Wissenschaft (10), 48–56.
Braitenberg V. (1986), Künstliche Wesen. Verhalten kybernetischer Vehikel. Vieweg, Wiesbaden.
Campenhausen C. von (1993), Die Sinne des Menschen. G. Thieme Verlag, Stuttgart.
Chalmers D. J. (1994), Explanatory gap; in "Toward a Scientific Basis for Consciousness" zit. nach Horgan 1994.
Churchland P. S. (1986), Neurophilosophy: Toward an Unified Science of the Mind-Brain. Bradford Books, MIT Press.
Churchland P. S. and Sejnowski T. J. (1992), The Computational Brain: Models and Methods on the Frontiers of Computational Neuroscience. Bradford Books, MIT Press.
Cowan J. (1986), Emerging Syntheses in Science, Proc. Founding Workshops of the Santa Fe Institute, (D. Pines, ed.), Santa Fe Institute, Santa Fe, New Mexico. pp. 184–193.
Creutzfeldt O. (1986), Gehirn und Geist. Nachdruck 1992. In: Rhythmogenesis in Neurons and Networks. Proc. 20th Göttingen Neurobiology Conference, (N. Elsner and D. W. Richter, eds.). G. Thieme Verlag. pp. 109–138.
Crick F. (1994), The Astonishing Hypothesis. The Scientific Search for the Soul. Simon & Schuster, London.
Crick F. and Koch C. (1990), Towards a neurobiological theory of consciousness. In: Seminars in the Neurosciences, Vol. **2**, 263–275.
Crick F. and Koch C. (1992), The problem of consciousness. Scientific American **267** (3), 111–117.
Damasio A. R. and Damasio H. (1992), Brain and language. Scientific American **267** (3), 62–71.
Damasio A. R., Damasio H., Tranel D. and Brandt J. P. (1990), Neural Regionalization of Knowledge Access: Preliminary Evidence. Cold Spring Harbor Symp. of Quant. Biol. LV, pp. 1039–1047.
Dawkins R. (1976), The Selfish Gene. Oxford University Press, Oxford.
Delbrück M. (1986), Wahrheit und Wirklichkeit. Über die Evolution des Erkennens. Rasch und Röhring Verlag, Hamburg, Zürich.
Dennett D. C. (1991), Consciousness Explained. The Penguin Press, West Drayton Middx.
Edelman G. M. (1992), Bright Air, Brilliant Fire. Basic Books (A. Lane, ed.). The Penguin Press, West Drayton Middx.
Egelhaaf M. and Borst A. (1993), A look into the cockpit of the fly: Visual orientation, algorithms, and identified neurons, J. Neuroscience **13**, 4563–4574.
Egelhaaf M., Hausen K., Reichardt W. and Wehrhahn C. (1988), Visual course control in flies relies on neuronal computation of object and background motion. TINS **11**, 351–358.
Eigen M. (1987), Stufen zum Leben. Piper Verlag.
Engel A. K., König P. and Singer W. (1993), Bildung repräsentativer Zustände im Gehirn. Spektrum der Wissenschaft **9**, 42–47.
Ermentrout G. B. and Cowan J. D. (1979), A mathematical theory of visual hallucination patterns. Biol. Cybernetics **34**, 137–150.

Fischbach G. (1992), Mind and brain. Scientific American **267** (3), 24–33.
Flanagan O. (1992), Consciousness reconsidered. MIT Press, New York.
Franceschini N., Pichon J. M. and Blanes C. (1992), From insect vision to robot vision. Phil. Trans. R. Soc. Lond. B **337**, 283–294.
Freeman W. J. (1988), Nonlinear neural dynamics in olfaction as a model for cognition. In: Dynamics of Sensory and Cognitive Processing by the Brain (E. Basar, ed.), Springer, Berlin, pp.19–29.
Freeman W. J. (1994), Chaotic behavior, hidden order; in "Toward a Scientific Basis for Consciousness" zit. nach Horgan 1994.
Gelernter D. (1992), Mirror World or: The Day Software Puts the Universe in a Shoebox. How it Will Happen, and what it Will Mean, Oxford University Press.
Gierer A. (1985), Die Physik, das Leben und die Seele. Piper Verlag.
Gödel K. (1931), Über formal unentscheidbare Sätze der Prinzipia Mathematica und verwandter Systeme, I. Monatshefte für Mathematik und Physik **38**, 173–198.
Goldman-Rakic P. S. (1992), Working memory and the mind. Scientific American **267** (3), 73–79.
Hassenstein B. and Reichardt W. (1956), Systemtheoretische Analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. Z. Naturforsch. **11b**, 513–524.
Hausen K. (1993), Decoding of retinal image flow in insects. In: Visual Motion and its Role in the Stabilization of Gaze, (F.A. Miles and J. Wallman, eds.). pp. 203–235.
Hebb D. O. (1949), The Organisation of Behavior. John Wiley, New York.
Hinton G. E. (1989), Connectionist learning procedures. Artificial Intelligence **40**, 185–234.
Hinton G. E. (1992), How neural networks learn from experience. Scientific American **267**, (3), 105–109.
Hofstadter D. R., (1979), Gödel, Escher, Bach: An Eternal Golden Braid. Harvester Press Ltd, Hassox, Sussex, GB.
Horgan J. (1994), Can science explain consciousness? Scientific American **271** (1), 72–78.
Kandel E. R. and Hawkins R. D. (1992), The biological basis of learning and individuality. Scientifc American, **267**, (3) 53–60.
Konishi M. (1993), Listening with two ears, Scientific American **268** (4), 34–41.
Konishi M., Takahashi T. T., Wagner H., Sullivan W. E. and Carr C. E. (1988), Neurophysiological and anatomical substrates of sound localization in the owl. In: Auditory Functions, Neurobiological Bases of Hearing, pp. 721–746.
Kurzweil R. (1990), Intelligent Machines. MIT Press, Cambridge, MA, pp. 117–118.
Livingstone M. S. and Hubel D. H. (1988), segregation of form, color, movement, and depth: Anatomy, physiology, and perception. Science **240**, 740–749.
Lorenz K. (1941), Kants Lehre vom Aphorismischen im Lichte gegenwärtiger Biologie. Blätter für Deutsche Philosophie **15**, 94–125.
Lorenz K. (1943), Die angeborenen Formen möglicher Erfahrung. Z. Tierpsychologie **5**, 235–409.
Lorenz K. (1978), Vergleichende Verhaltensforschung. Grundlagen der Ethologie. Springer, Wien, New York.
Mainzer K. (1994), Computer – Neue Flügel des Geistes? W. de Gruyter Berlin.
Martin N. and Franceschini N. (1994), Obstacle avoidance and speed control in a mobile vehicle equipped with a compound eye. In: Intelligent Vehicles (I. Masaki, ed.), MIT Press, Cambridge, MA, pp. 381–386.
Mayr E. (1979), Evolution und die Vielfalt des Lebens. Springer Verlag.
Männer R. and Lange R. (1994), Rechnen mit neuronalen Netzen. Physikalische Bläter **50**, 445–449 and 569–572.
Merzenich M. (1989), Representional plasticity in somatosensory and motorcortical fields. Biomedical Res. **10**, Suppl. 2, 85–87.
Mohr H. (1981), Biologische Erkenntnis. Teubner, Stuttgart.
Monod J. (1971), Zufall und Notwendigkeit. Piper Verlag.
Myers D. G. (1983), Social Psychology, McGraw-Hill, New York.
Nicholls J. G., Martin A. R. and Wallace B. G. (1992), From Neuron to Brain. Sinauer Ass., Inc., Sunderland, Mass., USA.
Penrose R. (1989), The Emperor's New Mind. Concerning Computers, Minds, and the Laws of Physics. Oxford University Press.
Penrose R. (1994), Shadows of the Mind: On Consciousness. Computation & the New Physics of the Mind, Oxford University Press, New York.
Poggio T. and Reichardt W. (1976), Visual control of orientation behavior in the fly. II. Towards the underlying neural interactions. Quart. Rev. Biophysics **9**, 377–438.
Quine W. V. O. (1968), Epistemology naturalized, Akten des 14. Int. Kongresses für Philosophie. Wien. Band **IV**, pp. 87–103.
Rasmussen S. (1994), Emergence; chaos and complexity; in "Toward a Scientific Basis for Consciousness" zit. nach Horgan 1994.
Reichardt W. (1982), Systems analytical behavioral research: as illustrated with the fly. In: Biophysics (W. Hoppe, W. Lohmann, H. Markl and H. Ziegler, eds.). Springer Verlag, pp. 811–820.
Reichardt W. (1986), Processing of optical information by the visual system of the fly. Vision Res. **26**. 113–126.
Reichardt W. (1987), Evaluation of optical motion information by movement detectors. J. comp. Physiol. A **161**, 533–547.
Reichardt W. and Poggio T. (1976), Visual control of orientation behavior in the fly. I. A quantitative analysis. Quart. Rev. Biophysics **9**, 311–375.
Reichardt W., Poggio T. and Hausen K. (1983), Figure-ground discrimination by relative movement in the visual system of the fly. II. Towards the neural circuitry. Biol. Cybern. Suppl. **46**, 1–30.
Richter M. (1993), Puran 2: Ein Zufallsgenerator zur Erzeugung von quasi-idealen Zufallszahlen aus elektronischem Rauschen. In: Messung, Modellierung und Bewertung von Rechen- und Kommunikationssyste-

men. (B. Walke and O. Spaniol, eds.), Verlag der Augustinusbuchhandlung, pp. 49–62.
Searle J. (1984), Minds, Brains, and Science. Cambridge: Harvard University Press.
Shepherd G. M. (1993), Neurobiologie. Springer Verlag.
Singer W. (1990), The formation of cooperative assemblies in the visual cortex. J. Exp. Biol. **153**, 177–197.
Singer W. (1994), The putative role of synchrony in cortical processing. Sensory Transduction. Proceedings of the 22nd Göttingen Neurobiology Conference 1994; Vol **I** (N. Elsner and H. Breer, eds.), Georg Thieme Verlag, Stuttgart, pp. 119–144.
Stieve H. (1990), Evolutionsprognose aus der Sicht eines Biologen. In: Evolution und Prognose, Alma Mater Aquensis. RWTH Aachen, pp. 157–167.
Strausfeld N. J. and Nässel P. R. (1981), Neuroarchitecture of brain regions that subserve the compound eyes of crustacea and insects. In: Handbook of Sensory Physiology, Vol. **7**: Comparative Physiology and Evolution of Vision in Invertebrates, Part B: Invertebrate Visual Centers and Behavior (H. Autrum, ed.), Springer Verlag, pp. 1-132.
Thompson R. F. (1993), The Brain: a Neuroscience Primer. W. H. Freeman & Co. N.Y.
Turing A. M. (1936), On computable numbers with an application to the Entscheidungsproblem. Proc.London Math.Soc. **42**, 230–265.
Turing A. M. (1950), Computing machinery and intelligence. Mind **59**, 433–460.
Vogel C. (1988), Gibt es eine natürliche Moral? Oder: wie widernatürlich ist unsere Ethik? In: Die Herausforderung der Evolutionsbiologie (H. Meier, ed.). Piper, München, pp. 193–219.
Vollmer G. (1975), Evolutionäre Erkenntnistheorie. S. Hirzel Verlag, Stuttgart.
Vollmer G. (1980), Evolutionäre Erkenntnistheorie und Leib-Seele-Problem. In: Wie entsteht der Geist? (W. Böhme, ed.), Herrenalber Texte, Karlsruhe, **23**, pp. 11–40.
Wagner H. (1994), Characteristic Delay: a new look at an old concept, Göttingen Neurobiology Report 1994. Proceedings of the 22nd Göttingen Neurobiology Conference 1994, Vol **II** (N. Elsner and H. Breer, eds.), Georg Thieme Verlag, Stuttgart, p. 390.
Wiesel T. N. (1982), The postnatal development of the visual cortex and the influence of environment. Nature **299**, 583–591.
Wittgenstein L. (1961), Tractatus logico-philosophicus. Deutsch-englische Ausgabe, New York.
Zeki S. (1992), The visual image in mind and brain. Scientific American **267** (3), 42–50.
Zeki S. (1993), A vision of the brain. Blackwell Scientific Publications.