3

* Captured in our Cockpits: Why Consciousness Evolved[1] \\\

3.1 INTRODUCTION AND OVERVIEW

The most central topic in epistemology, in the philosophy of mind and in the area of `animal philosophy' is the problem of consciousness. In recent years several attempts have been made to `explain' consciousness. Although many of the resulting models explain important aspects of consciousness, other fundamental aspects of consciousness are still neglected. Sometimes this is a consequence of presuppositions which cannot be critically assessed within the discipline from which the model stems. In this chapter I will claim that the different approaches can best be integrated within an evolutionary framework. Although many philosophers, neurobiologists and cognitive scientists think of themselves as evolutionists, they are not always aware of the consequences of Darwinism. Because they are relatively unfamiliar with the way in which evolution works, they are often too optimistic about the chances that particular kinds of minds can arise as a result of natural selection processes. It seems to me that it is this `semi-evolutionism' (or demi-Darwinism) that leads to pitfalls such as:

a. treating consciousness one-sidedly as a cognitive phenomenon (it could well be primarily a motivational phenomenon),
b. focussing one-sidedly on the neurophysiological level (forgetting that adaptations are likely to be emergent properties),
c.treating learning as the expression of the plasticity of the mind only (and forgetting that learning abilities are usually designed to acquire very specific skills and types of knowledge).

All these misunderstandings derive from an inability to see consciousness as an *adaptive phenomenon.* Consciousness is a good candidate for an `adaptation', because complex properties of organisms that require very specific combinations of genes are more likely to be products of variation and selection than of pure chance. Of course, consciousness could be an accidental by-product of another adaptive trait. This is unlikely, however, because the *subjective qualitative contents* of our experiences (e.g. pleasure or pain) are an integral part of our behavior. Negative experiences are linked to injury and malfunctioning and stimulate withdrawal, caution and care for oneself, whereas positive experiences are linked to health, growth and reproduction and stimulate activity, territorial expansion and reproduction (an argument going back to William James and Roy Wood Sellars). Some organs, like the heart and the intestines, seem to work independent of conscious control (being coordinated by the autonomous nervous system), whilst other organs seem to require conscious control at least during the obtaining of different skills (Baars, 1997). On the other hand, several defects have been discovered, like blindsight and the amnesic syndrome, in which conscious control of - and conscious access to - specific functions is lost without erasing those functions

[1] First published in: Ton Derksen (ed.), 1998, *The Promise of Evolutionary Epistemology*. Tilburg: Tilburg University Press, 101-130. Revised here as chapter 3 of *Why some apes became humans; competition, consciousness and culture,* Nijmegen: P.S. Press, 2002.

completely. However, nowadays such defects are no longer used to nourish the `epiphenomenalist suspicion' (Flanagan, 1992), because it is exactly the study of these defects that has enabled neuroscientists to compare cognitive processes which are consciously controlled with those which are not (Weiskrantz, 1997). As a result, students of brain and behavior are increasingly starting to consider the study of the evolutionary `Why?' of consciousness as a legitimate enterprise.

In this chapter I will defend the thesis that questions relating to the presence and nature of consciousness in animals and man can only be answered by interpreting consciousness as an adaptive phenomenon fully integrated into the behavioral repertoire with which a species has to cope within a particular environment. That implies that all theories which try to explain the neurophysiological `How?' question about consciousness should be integrated with the evolutionary `Why?' question, which should be asked from an all-encompassing evolutionary and ethological frame-work. I will try to show that in some current theories about consciousness the degree to which the brain is genetically prewired is still underestimated and I will propose that the phenomenon of radical subjectivity is best explained by cross-fertilizing evolutionary psychology with the theory of value-driven decision systems. Evolutionary psychology, building on selfish-gene theory, explains why consciousness is thoroughly perspectivistic and related to the interests of the individual or even of its genes. Value-driven decision-system theory, as developed by G.E. Pugh, explains the necessity of a decision system in which different behavioral options can be compared and `weighed'. I will try to show that consciousness constitutes a cognitive and motivational straitjacket which `subjectively forces' animals to act in an adaptive, and (inclusive) fitness maximizing, manner. `Innate structures of experience' are needed to supply values that enable animals to weigh different behavioral options: They supply heuristic approximations of the survival values of different experimental behavioral strategies (which may be explored during play). With a variation on a recent concept of Melzack, they may constitute a `genetically prewired neuromatrix for adaptive decisions'. The resulting evolutionary epistemology is somewhat more critical than German-Austrian evolutionary epistemology as it stresses the way in which information is subjectively transformed to narrow down the list of behavioral priorities for the organism. The model thus suggests a realistic and adaptive interpretation of qualia and stimulates the search for their behavioral correlates in animals other than man. At the end of the chapter, I will do an attempt to evaluate the evidence for animal awareness from this perspective.

What is consciousness?

Probably most readers would agree that consciousness can be viewed as a property of a living and working brain which enables the organism equipped with this brain to `live a life on its own', to experience its situation from its own perspective. Consciousness creates an unbridgeable gap between an individual that follows its own idiosyncratic course through its surroundings and all other organisms that can only guess at what the individual in question is intending. It thus constitutes a gap between inside and outside, between a domain of private and privileged access and a domain to which others may also have access.

Much philosophical confusion has arisen as a result of a tendency to reify both domains as different `substances', mind and body. On the one hand, a simplistic identification of those `substances' will not suffice, because much more activity goes on in our bodies and brains than we are consciously aware of. The phenomenon of death shows that consciousness is much more a property of specific bodies and brains than that it can be identified with those bodies/brains. On the other hand, most philosophers don't want to go back to an attempt to view the mind as a kind of transbiological phenomenon that `hovers' above the brain and makes contact with it via, for example, the pineal gland (as Descartes thought). I defend the view that consciousness is *an adaptive function, embodied in a particular brain organization, that enables certain organisms to interact effectively with their environments in such a way that they are able to `calculate' and pursue their own unique interests within that environment from their own unique perspective.* Consciousness is in this view completely linked to making very complex adaptive choices. The problem is, of course, that we all know that many, many conscious choices are not adaptive. The argument that "nature is not perfect" and that man is probably not optimally adapted to its current environment sounds as a bad excuse, similar to the `justification of God' or theodicee within the theistic world-view , but may nevertheless offer the a better solution than throwing out the baby (of adaptation) with the bathwater (of perfectness).

Before I start defending the evolutionary approach, however, let us try to agree on a number of `phenomenological' issues that have to be dealt with within a complete theory of consciousness. Most readers will probably agree that an adequate theory of consciousness has to explain some of the following properties on which philosophers have dwelled for centuries:

- Consciousness is characterized by *qualia*, the Latin word that was originally selected by Cicero to translate Plato's `poiotês', `of-what-kind-ness' (Barlow, cited in Baars, 1997: 82). Originally the word referred to characteristics of objects, but it is used currently to refer to our unique subjective experiences of such characteristics. There is a difference between the different wavelengths of visible light and the way an organism *experiences* them. Additionally, one cannot describe one's pleasures and pains simply by `measuring' their values on some quantative scale. The problem is that qualia are only perceptible from a first-person perspective and that one cannot compare first-person perspectives objectively: Long conversations with much empathic projection are the only means of obtaining glimpses of the internal states of other persons.
- Consciousness is often linked to *intentionality*. Intentionality derives from the Latin word *intendere* which seem to have arisen in the context of fighting with bow and arrow and means: stretching, aiming at, threatening with, etc.. One could translate it as `aboutness', claiming that we are always conscious `about' something (Brentano's thesis). Often objects or relations surrounding an organism are somehow `represented' in its mind. This `representation' should not be seen as a mind-copy of the object, but should be seen as the organism's *cognitive gateway* to that object, the `sign' or `access-code' which enables the organism to deal with the object and to learn more about it.
- Consciousness seems, to some extent, linked to *`control'* and *volition*. In the cognitive realm that means that it may result in the control of attention; for the organism as a whole this means that consciousness seems to be linked to volition and action. Traditionally the so-called freedom of the will was often opposed to determinism, but within a naturalistic frame-work it should only be opposed to a narrow physical, chemical or genetic determinism and the will can be called free if it is determined by a weighing of the desirability of different scenarios (`soft determinism'). A transcendental free will, which does not somehow reside in the unique properties of some brains, cannot be the object of scientific thinking.

- As shown by many writers, including Augustine, James, Joyce and Proust, consciousness has the character of a *stream*. In fact a lot of small brooks lead into it and it shifts its course continually, which makes its behavior hard to predict. It certainly does not have the character of a series of logical deductions from observations and it often enables us to be occupied with different problems at the same time, shifting our attention from one to the other continually. (If we are unable to retrieve a name, for example, it often ascends later unexpectedly, as if a search-program had been working backstage all the time).
- Consciousness, at least in human beings, seems to result in *self-consciousness*, an awareness of an own unique self which is different, and to some extent isolated, from the rest of the world. This self offers both the context of the series of decisions which emanate from it as it is, to some extent, changed and constructed by them.

3.2 CONNECTIONISM, CORTICOTHALAMIC LOOPS AND FORTY-HERTZ OSCILLATIONS

Consciousness is currently more intensively studied than ever before. Over the last ten years a series of new discoveries have been made which will probably have to be integrated in any mature theory about consciousness. As always, such discoveries will sometimes lead to enthusiastic and exaggerated theoretical claims. Before introducing the perspective taken in this chapter (3.3), I will briefly discuss some current perspectives on consciousness. I will look whether these theories, developed by people who mostly think of themselves as darwinians, really do take the lessons of the `second darwinian revolution' (Cziko, 1995) to their heart. In part 3.4 I will show, then, how the valuable elements of these perspectives can be integrated into the theoretical framework defended here.

(Negative) feedback control theory. Already William James noted that organisms differ from other natural systems in that they produce consistent ends by variable means. In the twentieth century Norbert Wiener showed that certain neuromuscular `reflex arcs' are based on the principle of negative feedback. Gradually it was realized that his cybernetics could explain how organisms are able to maintain certain goal states independent of fluctuations in the environment. In that, they were like thermostats, but thermostats that continually have to work to stay balanced.

In recent years William Powers and the `Control Systems Group' have proposed that negative feedback theory in the end can explain most design features of the brain. The psyche is essentially seen as device that enables organisms to stay directed to their goals, independent from external influences. `A control system, properly organized for its environment, will produce whatever output is required in order to achieve a constant sensed result, even in the presence of unpredictable disturbances' (Powers, 1989: 77).

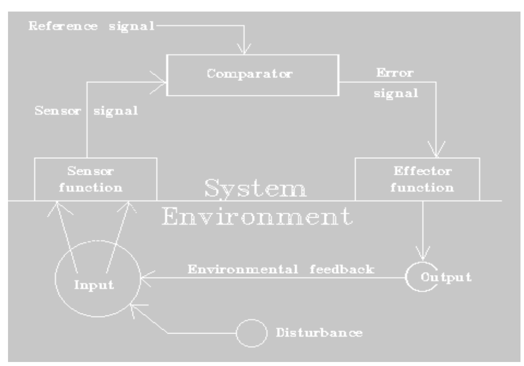

Figure 3-1. Control-system unit according to Powers. The system corrects differences between sensor signal and reference signal by producing an output which influences the input.

In the model of Powers (see figure 3-1)

a disturbance of the environment affects an organism via a sensor signal which in a `comparator' is compared to a `reference signal' and leads via an `error signal' to behavior that restores or changes the environment in the direction that is desired. If there is a place for consciousness and qualia in this model they should probably be placed somewhere between the reference and the error signal. Apparently consciousness should be viewed as the way in which nature enables us stay in a `dynamic equilibrium' and to be faithful to ourselves in a changing environment. As the writers of the Control Group's `Introduction to Psychology put it:

> 'The `stream of consciousness' seems to be the play of perceptions in systems which are making adjustments, or, in other words, the focus of attention follows the largest-magnitude error signals anywhere in one's systems' (Robertson & Powers, 1990: 210).

Thus, control theory interprets the brain as a complex `homeostatic' device which has to keep the organism in control of its environment. According to the Control Group `the overall purpose for which an organism controls its environment is to maintain and optimize its existence' (idem: 86).

Is this theory sophisticated enough to explain the nature and evolutionary necessity of consciousness? It is certainly an improvement compared to the behavioristic stimulus-response models, but seems still much to simplistic to explain consciousness. Organisms do much more than maintaining and optimizing their existence by reacting to environmental disturbances. At least those that have to move and possess brains are actively pursuing a whole list of goals, including different types of food, safety, sleeping sites, sexual partners, the well-being of their brood, etc.. In order to be able to reach those goals, many intelligent organisms have to make choices, to look forward and to plan. This is much more than just `correcting error signals'. Despite its name the `Control View' seems only to be about `correction': real control also includes the ability to remain faithful to a set of goals in a completely unexpected situation. Real control implies the cognitive assessment of such a new situation and the ability to react appropriately, despite the complexity of the situation. The Control View is much too general to be particularly enlightening with respect to this ability and, therefore, to consciousness.

Churchland and Connectionism. The study of the properties of neural networks is as old as cybernetics. In 1959 Rosenblatt proposed the idea of a *perceptron*, consisting of a series of input units linked to output units via adjustable connections (Harth, 1993: 55). During a series of instructional steps the strengths of the connecting signals could be adjusted so that the perceptron could be `trained' to link complex input patterns with specific outputs. As a result of the success of serial computers, the concept of neural networks was overtaken for some decades by the idea of software `computer programs' that can be run on any hardware digital computer.

However, during the eighties the concept of neural networks won new interest as a result of the discovery of new ways to improve the performances of the old `perceptron'. A layer of `hidden units' was inserted between the input and the output layers and new rules were introduced to make it possible for each unit to adjust its `synaptic weight' effectively to that of the adjacent units, if an undesired output was produced. As a result the new networks could be `trained' to make certain distinctions simply by instructing them to adjust their synaptic weights according to specially designed rules if they were giving the wrong output.

As we saw in chapter 2, the new networks proved very successful in `learning' to `recognize' complex input patterns and to link them to particular distinctions and even decisions (Churchland, 1989; see figure 2.1 to 2.3). This was the more surprising as the only way in which the `information' was stored was a distribution of `synaptic weights' in a neural network. The particular distribution of `synaptic weights' that enables the recognition, or `activation', of particular patterns is called a prototypical `vector'. Parallel distributed processing showed very clearly how real neural networks could *learn* to make certain distinctions by trial, error and effective adjustment rules. (In nature, of course, these adjustment rules themselves have to have their origin as a result of natural variation and selection.) As such they show how conceptualization is possible within a material device.

Initially, Patricia and Paul Churchland, who have become the major philosophical defenders of connectionism, did not give an explanation of consciousness. They even tended to evade the problem and suggested that a first-person account is not characterized by specific unique contents, like qualia, but only by a specific and very inadequate conceptualization of the things that go on in a neural network: `folk psychology' (a concept showing the influence of Wilfrid Sellars who was one of the first to state clearly that even the way in which we observe ourselves is theory-laden; Sellars, 1963). In the end Paul Churchland hoped that we could learn to use a more appropriate language to reveal the idiosyncratic caprices of our neural networks to each other.

> Given a deep and practiced familiarity with the developing idioms of cognitive neurobiology, we might learn to discriminate by introspection the coding vectors in our internal axonal pathways, the activation patterns across salient neural populations, and myriad other things besides (1989: 75).

From this sentence it is clear that Churchland does not see qualia as constituting an independent ontological level, as phenomena that are somehow part of the way in which our mind works. As a typical epiphenomenalist he tends to `quine qualia' (to quine means `to deny resolutely the existence of seemingly undeniable phenomena', Dennett, 1990; Flanagan, 1992). Qualia are perceived as merely inadequate descriptions of the complex neurocomputational processes that go on in our minds and `nothing more'. Churchland hopes that these inadequate descriptions can be replaced by descriptions in a more scientific terminology, just as water `really' is H_2O and red is `really' electromagnetic radiation with a wavelength of somewhat less than 10^{-6} metres. At least until his 1989 work it seemed to be his hope that an adequate scientific terminology could completely replace the first-person perspective: We would be able to refer to our emotional states not in terms of such crude categories as `love', `pain', `hunger' and `hope', but in terms of specific neural fibers giving signals from one group of neurons to another group. This we could call the `Quine qualia' or QQ-thesis. As will be seen later in this paper, the hope that this thesis may be right results from both a misunderstanding about qualia and about introspection.

In a later work (1995), Churchland has placed his bets on `recurrent networks' in which `recurrent (descending) pathways' are projected back from the output level to intermediate levels. Such recurrent network display properties which one does not find in simple `feedforward networks'. For example, the `recurrent pathways' can bias the network in such a way that particular prototypes are activated instead of others. Recurrent networks are not continually dependent on new input: They can generate complex sequences of activation `vectors' all by themselves. Recurrent networks are also not only able to pattern recognition, as feedforward networks do, but they are also able to recognize and represent prototypical processes which are extended in time. They thus enable one to recognize causal sequences and even to predict the future to some extent. Thus, the dimension of time is added to the neural network as a result of descending pathways, which add information about the network's past activities to its current activities.

Following suggestions made by Francis Crick (1995; Crick & Koch, 1992) and others, Churchland proposes in his 1995-book that the phenomenon of consciousness should be explained as the result of the ascending and descending axonal pathways that connect the cerebral cortex with the intralaminar nucleus of the thalamus - in short, as a `network property' of a large-scale *recurrent* neural network. The difference between sleeping, dreaming and waking can be explained as a result of the ways in which the intralaminar nuclei (clumps of nerve cells inside the two thalami, a loss of which produces an irreversible coma) can stop initiating 40 Hertz oscillations (see below) and can shut down the recurrent pathways back to the cortex. The phenomenon of attention can be explained as the shifting pre-activations of particular neuronal layers in the cortex in such a way that particular prototypes have a bigger chance of becoming part of the global oscillations (if you are looking for your cat, all kinds of cat-features and cat-related associations are pre-activated).

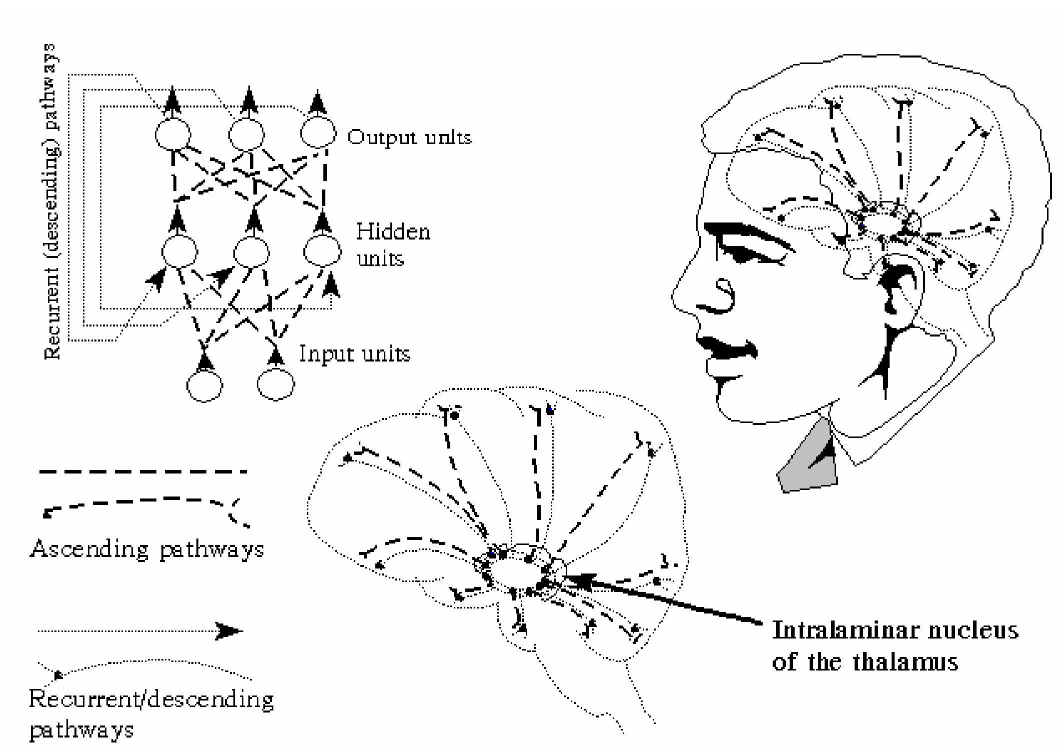

Figure 3-2. Because it communicates back and forth with the cerebral cortex, the intralaminar nucleus is supposed to `carry out' the function of consciousness (Adapted from Churchland and Llanás).

It remains to be seen, however, whether even this attempt at `reducing' consciousness

(Churchland keeps flirting with eliminative materialism) is not somewhat too simplistic. Connectionism seems to be able to explain abstraction, pattern recognition and even cognitive biases, but is by itself unable to explain why organisms have internal goals, subjective experiences, intentionality and self-consciousness. Given the fact that Churchland has committed himself to the defense of the almost infinite *plasticity* of the human mind, I do not see how he can bridge the gap between the collection of very general cognitive mechanisms that he is able to explain and the behavior of real organisms that have to pursue very specific goals in an often unfriendly environment. If Churchland continues to defend the plasticity of the human mind - and does not tone down on this point - he will also never come to understand why particular qualia are coupled to particular situations. If our evolutionary history had not `programmed' us otherwise, the act of eating could go hand in hand with nausea, sex could be horrible and drowning could feel absolutely fabulous, etc. The `meaning' of particular qualia, the specific connection between a phenomenological quality and the series of objective situations which it evokes, can only be understood in the light of their evolution.

Edelman, Crick, creative loops and the forty-Hertz hypothesis. Churchland is not the only one to have come up with recurrent networks. Many different authors have proposed the idea that the essence of consciousness lies in some kind of loop. Harth has proposed that positive feedback loops (reinforcement) in the form of `creative loops' play an important role in some kind of cognitive `bootstrapping' which eventually leads to consciousness. The idea is that `particular fluctuations are amplified selectively so that features not initially present at the input may be *generated* in a *bootstrap* fashion' (Harth, 1995: 71; see also figure 2.5, 2.6, 2.10 which were all derived from or inspired by Harth). The resulting images may then be compared with the input again and so forth. `Consciousness, which arises in this self-referent process, not only unifies the immediate sensory messages but also becomes the joiner of everything around us, past, present, and future' (Harth, 1995: 144).

Loops in the form of `reentry' also play a significant role in the `theory of neural group selection' of Gerald Edelman (1992). Even on the level of what he calls `primary consciousness' (as distinguished from `higher-order consciousness') a `reentrant' loop is crucial. Edelman thinks that primary consciousness is already based on a residue of past interactions with the environment in which value-laden categories have arisen. Primary consciousness arises at the moment that these value-laden categories activate (re-enter) current information processing of new sensory input. Therefore Edelman speaks of primary consciousness as a `remembered present' based on a circuit that `allows for continual reentrant signaling between the value-category memory and the ongoing global mappings that are concerned with perceptual categorization in real time' (Edelman, 1992: 119). Physically he locates the `value-category memory' in the frontal, temporal and parietal cortex which interacts with deeper parts of the brain like the brain stem and hypothalamus via the hippocampus, amygdala and septum. The input of `world signals' is processed in the primary and secondary cortex.

According to Edelman, the circuit in which primary consciousness emerges also creates the possibility of `higher-order consciousness' in animals with language or language-like abilities. Between the primary and secondary cortex in which sensory input is processed and the value-category memory, Broca's and Wernicke's areas are placed. These allow for a new kind of conceptual memory which can probably compress more data in a much more efficient way. This leads `via semantic bootstrapping' to a `conceptual explosion', as a result of which `the self, the past, and the future can be connected to primary consciousness'. The animal becomes aware of the fact that it is conscious: Consciousness of consciousness becomes possible.

Many different authors have observed that consciousness somehow seems to solve a `binding problem'. As Churchland repeatedly stresses, the brain is a massive *parallel* `computer' and somehow all lines of independent information processing have to be bound together to be coordinated. Both Rodolfo Llinás and Francis Crick (together with Koch) have developed the theory (already mentioned above), inspired by research and suggestions of many other neuroscientists, that consciousness emerges from the correlated firing of a large number of neurons at an oscillation of 40 cycles a second (40 Hertz) that has been found to be characteristic of the whole cortex and of the intralaminar nucleus during waking hours (Baars, 1997; Crick & Koch, 1992; Churchland, 1995). According to Crick (1995) spikes arriving simultaneously at a neuron will produce a larger effect than the same number of spikes arriving at different times. Consciousness thus plays the role of integrating information by forcing independent thoughts to `dance the same tango'. Crick points to the thalamus as the brain's `organ of attention' which directs its `spotlight' to neural areas where a lot of information has to be integrated by imposing its 40 Hertz `beat' upon them.

All this is certainly very interesting and might well be true. At the same time it does not yet explain why we need a private domain in which we are emotionally and cognitively isolated from the rest of the world. It remains to be seen whether neuropsychology can answer that question.

Baars and his global workspace theory. The forty-Hertz hypothesis is based on a notion of consciousness as an *integration* of information that would otherwise be independent. The idea that it is the function of consciousness to coordinate the information processing that goes on in many different parts of the brain is far from new. Ten years ago Bernard Baars had already developed his `Global Workspace theory' in which consciousness is compared to a theater in which a lot happens behind the scenes, but in which everything is focused on what happens `on-stage'. Automatic processes and routines happen everywhere in the brain, but the moment new habits have to be learned and behavior has to be changed all relevant information has to be recruited from all of these dispersed brain regions. Consciousness should be seen as a field of heightened neuronal activity which enables problems to be solved which can only be solved by linking brain regions which lie far apart. In the words of Baars:

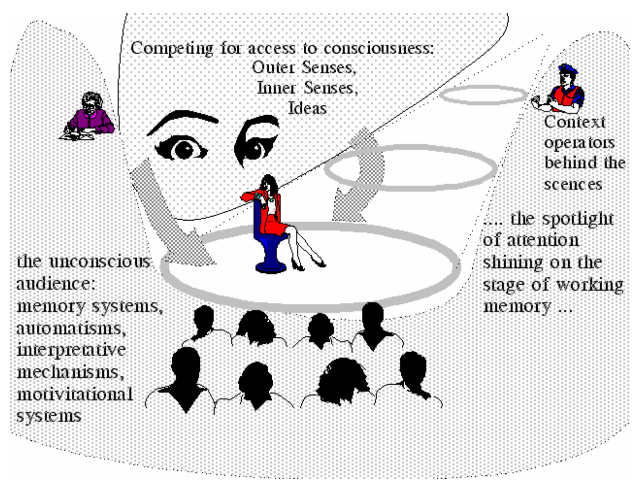

Figure 3-3. Baars' theater metaphor for conscious experience. Different inputs *converge* into conscious experience, yet *diverge* again to all kinds of intelligent unconscious mechanisms.

It seems that the single most prominent function of consciousness is to

> increase access between otherwise separate sources of information (Baars, 1997: 162).

Baars himself thinks his `theater model' emerges naturally from all recent developments of neuroscience, including the discovery of the forty Hertz oscillations in the cortex and extralaminar nuclei. The theater metaphor, although very enlightening in some respects, also has some disadvantages, however. As Baars would probably be the first to admit, what happens `on-stage' will often result in action: A parliament metaphor would be better in that respect (a parliament being a theater in which the actors at least think they have some power). Only in this respect is Baars' theory still somewhat too much based on the presupposition of a purely *cognitive* function of consciousness.

Interim conclusion. Various different theories about consciousness and the brain will probably prove to be mutually compatible as they are all portraits of the brain from different perspectives, on different scales and with a different degree of resolution. Connectionism explains how neural networks can store `concepts' and can even `learn' by continually readjusting their synaptic weights. Recurrent networks are probably the best technological parallel to simple systems of neurons. Loop theories refer more specifically to the possible physical properties of the brain that allow it to display consciousness: Consciousness is identified with the reverberation of an electronical tornado between thalamus and cortex. The forty Hertz hypothesis gives at least the beginning of a solution to the `binding problem' and is compatible with the Global Workspace theory which stresses the role of consciousness in integrating a great deal of information from different parts of the brain.

Some of the above mentioned properties of consciousness are explained by all these models. For example, the theory about recurrent networks can explain why the activities of the brain are to some extent independent of the environment. Connectionism can explain why concepts in our minds are not `defined' clearly and do not correspond exactly with an aggregate of `meaning atoms' within clear cut boundaries. Loop theories can explain why consciousness is more a `field' or `stream' than a calculator or a serial computer which essentially thinks in the way in which only a hyperintelligent mathematician or rationalistic philosopher can talk, occasionally.

Yet, something is still missing. To my mind, at least, all of these theories still do not explain the most essential property of consciousness: the radical *private* nature of consciousness and the way it creates a gap between `inside' and `outside' the organism's unique `point of view'. Workspace theory is, in principle, compatible with a more radical Darwinian approach, but a much more sophisticated theory about sensation is needed to account for its radical, unavoidable, qualitative and private nature.

3.3 CONSCIOUSNESS AS AN ADAPTIVE PHENOMENON

Sometimes it helps to step back a little to get a wider perspective (or to replace your close-up lens with a wide-angle). It is by no means self-evident that consciousness is better understood by observing it as closely as possible, be it from the inside, as phenomenologists have done, or from the outside, as physiologists do. One way of stepping back from both perspectives is to see things within an evolutionary frame-work. Perhaps consciousness should be viewed as a biological phenomenon that we can understand better *after* we have understood more about evolution in general.

Selfish gene theory and evolutionary psychology. In my view, it is especially the `selfish gene'- approach that is enlightening in this context. Because I will give a historical analysis of the sociobiological `bottom-up' perspective in chapter 4, I can be short about it here. At this point, I only want to stress that the view of the individual organism as `DNA's way of making more DNA' (Wilson, [1975] 1980) or as the `survival machine' that is designed to help replicate the `selfish genes' that have built it (Dawkins, 1976) are to be understood as short and lively ways of stressing the `bottom-up' architecture of nature in which higher and more emergent entities like societies should be understood at their proper level. It is not always the smaller building block that makes the bigger entity of which it forms a part understandable, however. Rather, we should try to locate those building blocks that have been chiseled out by nature itself, in the form of natural selection or differential reproduction. Only if we understand the levels at which selection in nature really works can we understand the elements of `design' inherent in the features of living objects. Although selection may have worked on many levels, it is especially the individual as a fortress of an alliance of `selfish genes', which often forms an integrated building block of self-interest in nature. It seems to me that this view throws light on the phenomenon of consciousness. Why?

First, it explains why the knowledge-apparatus has evolved to serve individual organisms in the first place and why the information in consciousness is stored in a thoroughly perspectivistic way. As the cybernetic resources of information processing systems are always limited, only relevant information can be processed. Information is relevant insofar as it leads to survival (of genes and their temporary `survival machine' or `vehicle') and for this reason information concerning the direct environment is valued above anything else. This information does not need to be `objective', but *should* be perspectivistic, because it should also entail information about the current position of the organism (a lion is conceived quite differently by its fleas or by a candidate-prey, like a wildebeest).

Secondly, for this reason the way in which this information `touches' the organism should be `value laden' and relate directly to its interests: Only in this way can the organism perceive something as `threatening', `dangerous', `attractive', etc. *So, perspectivistic, value-laden information processing is to be expected according to a `bottom-up' selfish gene approach to organisms. Even in a hypersocial environment it is important that an individual keeps an eye on its own interests and, therefore, gathers information from its own perspective.* Of course, even in this situation most of the information relevant to an organism concerns its own body, its position and the availability of dangers, food and potential partners in its direct environment.

Thirdly, talking about `survival machines' may help us to remember why minds are not in the first place `information processing tissues', but biological `decision centers' and why the information used by these `decision centers' need not be represented completely to the central parts of it. It is to be expected that a survey is somehow made of the information most relevant to the decisions that have to be made: The different options have to be clear, but it does not need to be clear how a process of data-compression led to these options. It is to be expected, therefore, that an individual survival machine somehow, somewhere, keeps an `overview' of the most crucial information relevant to the most necessary and inevitable decisions. Consciousness could serve this function. It integrates information relevant for behavioral options. So conceived, *consciousness can be viewed as a way in which a survival machine is given an overview of information that is immediately relevant for its behavioral options* (one notes here the affinity with Global Workspace theory; I will come with my own metaphor later).

Finally, the selfish gene approach may also help to explain why consciousness, at least in many mammals, is a private phenomenon, even though these species display a relatively sophisticated level of mind-reading. Why would the information that is used to steer the organism be available primarily to the organism itself? This way of organizing things would be very awkward if individual organisms behaved for the good of the species or even for the good of the group. If it is the good of the group that matters, then not every individual needs to collect all necessary information for itself: Ant colonies, with a high degree of kinship, can also react collectively to new situations and even `learn' collectively (Gordon, 1995).

Even if it was necessary to integrate all information in each individual, individuals could be like open books to each other and inform one another continually about all new mutually relevant information. Clearly, this does not happen in many mammalian societies, including humankind. Our societies are not `superorganic', `top-down' superorganisms in which individuals almost always form cooperating `cells', but a kind of `bottom-up' emergent group-contract to which everyone agrees as long it is in his/her interest. The possibility of parasitism looms everywhere and all financial and informational systems have to be secured against fraud by elaborate procedures. Any personal information can be misused at any time and this is probably the reason why we keep many things secret, even from ourselves.

Thus the selfish gene approach may explain why the information in consciousness is always perspectivistic, up-to-date and related to the interests of the conscious individual and his/her genes, the `first-person' of consciousness. However, we do not know yet why - often - `survival machines' are not robotic automata and - often - seem to have been designed to *experience* the information that they have to process.

Value-driven decision system-theory. Information in itself is neutral. Something is simply true or not true and by means of logic one can deduce other truths from it. A survival machine obviously does not need information for the sake of information and, as we have seen, it should therefore restrict itself to relevant information. How, then, should it select relevant information? Obviously one way is by not having unnecessary sense-organs. But probably even the most necessary sense-organs create a lot of noise. Above that, information that comes in from different senses somehow has to be matched together in a picture of the world on the basis of which the organism can `compute' its priorities, the `vector' of its will and/or its intended course through the outside world.

How should it compute such decisions? Selfish gene theory suggests that the central decision system within an organism will behave like a parliament of representatives, but that does not exclude *one* resulting teleonomic `vector'. In most cases this vector is directed to a maximum of offspring that can be raised successfully given the current circumstances (in albatrosses and chimpanzees this does not lead to a high number of offspring). However, the organism has to achieve this goal in a completely unpredictable environment. Of course, it is possible to instruct its DNA with an amount of potential strategies and with a series of routines that enables it to discover its environment and then to choose the most adequate behavioral strategy. Obviously, an enormous brain would be needed if it had to be programmed with all possible alternative strategies. The chances would also be high that the strategy triggered by a specific set of stimuli will be slightly inadequate in a new environment; the consequences of such small inadequacies could be disastrous.

Control or feedback theory gives one possible solution to this problem of an unpredictable environment. One specific control variable is specified that has to be kept constant at some specified control level. Compensating actions can be specified too, even in complicated decision tables, so that the organism knows what to do when the variables move outside the specified control range (e.g. body temperature can be corrected by increasing metabolism). A whole hierarchy of control systems can together create a sophisticated control system in which several partly independent feedback loops are intertwined (Powers, 1989).

Within this approach the problem of the unpredictable environment is not completely solved, however. In many planning processes the series of decisions that have to be made sequentially is so complex that it cannot be derived simply from a series of control variables and a related decision table. Sometimes even the control variables can no longer be kept constant, and have to be adjusted. The only adequate decision table for such a system would consist of an almost infinite branching tree of decision tables in which the possible alternative consequences of certain decisions in certain environments would have to be anticipated and in which an adequate way of keeping the relevant control variable within an acceptable range would have to be specified in advance. If these decision tables had to be `installed' in actual organisms, they would need enormous brains.

George Edgin Pugh describes in his *The Biological Origin of Human Values* (1978) how he was working for the U.S. Defense Department in the late 1950s and the early 1960s to develop a computer system for the automatic development of bomber flight plans. It proved impossible to instruct the program simply with the series of rules of thumb that the military experts had provided, because they proved to be contradictory and inconsistent for numerous situations. This had not been a real problem for the military commanders in the past, because they could always fall back on their common sense if the rules proved to be in conflict or inapplicable. If the inconsistencies were removed enormous lists of exceptions and new rules emerged and the resulting decision process became even more complex without increasing effectiveness.

The only way in which it proved to be possible to enable a computer program to devise optimal bomber fighter flight routes and schedules was to enable the program to list a large number of alternatives and to score them by means of heuristic values assigned to both the aircraft and crew and the targets destroyed (multiplied both by the probability of recovery and destruction, respectively). The introduction of these heuristic values made it no longer necessary to think through every contingency in detail and the system could even devise new alternatives, the possibility of which the planners had not foreseen. Because the program would always note the possibility of disastrous consequences if the values were scaled appropriately it was possible to let it explore large numbers of alternatives and to find optimal solutions which would not otherwise have occurred to a human being. (Such a program can probably also be produced within a connectionistic architecture. One can `train' such a network by feeding it with a large number of prototypical input-output relationships; perhaps values arise the moment output is given in broad classes of behavior).

The importance of Pugh's work lies in his use of the concept of such so-called *value-driven decision systems* to throw light on biological intelligence. He notes that there is an obvious parallel between the problems of too-simplistic artificial and organic `decision systems'. Light-seeking moths that fly into flames, light-seeking flies and wasps that keep bumping against windows and hedge-hogs that roll up in front of cars all show the limits of a too-limited collection of preprogrammed action patterns. Like a human designer, evolution or the virtual `evolutionary designer' (a concept used by Pugh to make it easier to speak about evolution's `design problems') is almost certainly unable to foresee all the situations in which its designs will be put to the test. The only way in which flexibility can be built into a system is by giving up the idea of complete preprogramming and by introducing representations of the world into the system combined either with a notion of ultimate goals or with heuristic values.

However, building a notion of ultimate goals into a decision system will almost certainly lead to overcharging its cybernetic resources. An obvious parallel is offered by the game of chess. Ultimately, the only real objective in a game of chess is to win i.e. to achieve a checkmate. Yet, experienced chess-players and chess computers will assign values to specific pieces, like the queen, and to certain favorable or dangerous configurations. These values are a `surrogate' for the real goals. The perfect player would not need them: Such a player would `simply' compute every possible consequence of every individual move up to the last move of the game each time that it is his/her turn. Such a perfect player would need almost infinite cybernetic resources, however, because the amount of possible reactions on the part of the opponent multiplied by the amount of possible next moves multiplied by the amount of next reactions of the opponent multiplied by the amount of the amount of possibilities thereupon, etcetera, will be enormous.

Obviously the game of gene survival in a complex environment has still more possible moves and at least some players have more turns than in the game of chess. The ultimate goal can be described as the maximization of gene replication (Pugh himself still thinks in `survival of the species' terminology: *op. cit.* 73), but biological decision systems would need very large brains and these would be very slow if they had to calculate the consequences of each possible move in function of this goal each time they had to make a decision. In addition, it is not clear how a process of natural selection which has to lead to at least some adapted individuals in each generation, could result in the evolution of such decision systems. Such systems would obviously be outcompeted by systems that were less wise, but smaller and faster. In fact, in present-day warm-blooded animals the process of natural selection seems to have led mostly to relatively small, compact and very efficient decision systems in which the ultimate goals of gene survival are represented by a series of heuristic values assigned to such different topics as food, sex, predators, safety, comfort, good company, pleasurable and repulsive smells, hygiene, etc.. If evolution ever had a choice between the representation of ultimate goals and the adjudication of `surrogate' values, it is clear that it has made its choice for the latter.

How, then, are these `surrogate' values represented? In the case of the `human decision system', Pugh makes a distinction between primary or innate and secondary or derived values. The latter are the values of everyday conversation, the values that many people think to be `culturally determined', because they are slightly different in different cultures and in different times. According to Pugh the *primary values correspond to the elementary valuative sensations of human consciousness*, like discomfort/comfort, pain/pleasure, bad/good taste, bad/good smell, sorrow/joy, shame/pride, fear, anger, hunger, thirst, itch. They are to a large extent innate and they represent decision criteria built into our brains as a result of random mutations and the survival of the fittest. Pugh thinks of them as built-in valuative sensations that are `a result of physical linkages in the neurons of the brain, which are inherited in exactly the same way as other physical characteristics'.

Pugh's belief in the innateness of a collection of fundamental human values is not a result of dogmatism. His experience as a programmer using value-driven decision systems has taught him that a system begins to behave much more adequately if the number of preprogrammed heuristic values increases (Pugh, 1978: 66). His most important example in this respect is the development of an automatic `student assignment system' which had to achieve a maximum of racial balance in public schools with a minimum of busing for individual students. (From his references we may conclude that he worked for two years, 1971 and 1972, on this problem or its reconstruction.) The development of a value structure that could produce an optimum solution started with a simple concept of a value structure: A positive score was given according to the level of desegregation achieved and penalties were given for every child that had to ride more than thirty minutes. This simple value structure had to be refined, however, in six steps, before a more intuitively attractive distribution of students arose: small extra penalties had to be given for each student that had to use a bus, for each minute at the bus, a nonlinear travel penalty had to be introduced, travel penalties had to be increased and an extra penalty had to be introduced to discourage different school assignments for the same neighborhood area. Pugh concluded that the introduction of extra values, sometimes negative ones, is essential for the fine-tuning of a sophisticated decision-process in a complicated real-life situation. The parallel to the evolution of biological decision systems is obvious: Multicomponent value structures simply have to be expected given enough time and ecological variation for experimentation. This applies especially to species with a relatively small number of offspring living in relatively complicated ecological and social situations, because the value of each individual `vehicle' is relatively high in this case and it may well be worth the trouble to equip it with a talent for sensible decisions[2].

One obvious advantage of an innate value system is that the organism that is equipped with it does not have to learn as much, or at least knows *what* it has to learn (it knows the value of particular information). However, there is another fundamental reason why the value structure has to be innate and irrational, and not subject to change on the basis of rational thought. The survival machine has to serve evolutionary objectives that it does not need to understand. If it were allowed to adjust its own primary values it might start to drop all kinds of penalties for which it does not understand the reason. It could even assign positive values to poisonous food, dangerous situations, and non-adaptive behavior. Narcissism, escapism and suicide could result. It is clear that the ability to change the own primary value structure is in contradiction with the very idea of a value-driven decision system which has to serve the objectives of a designer beyond the system itself.

That does not mean that in both artificial and biological decision systems there is no room for the adjustment of particular *secondary* decision criteria which might enable the system to learn from experience (Pugh, 1978: 32). In the case of humans these might be inherited culturally. However, such secondary values always have to be evaluated against the primary values. Perhaps the cultural success of particular secondary value systems in our species depends largely on their effectiveness in adapting characteristics of the primary value structure to a particular ecological and social niche and the life style required by the economic possibilities that it offers.

All this means that our minds are less plastic than philosophers such as the Churchlands suspect and that the information that ends up on our desk, the desk beneath our

[2] Mammals are a group whose number of offspring is naturally limited as a result of their system of internal hatching and intensive parental care (they are relatively `K selected'; see chapter 6). Additionally, the size of their brains is not limited by their respiratory system, by an exoskeleton or by the necessity to fly.

skull, is already censored and colored. On the one hand the primary values seem to be innate, on the other hand they seem to be conscious. That means that consciousness is not as related to plasticity as is sometimes thought and that, to the contrary, *consciousness is an inborn mechanism of weighing the survival value of incoming information.* Consciousness enables us to be `plastic' only by having a particular structure. *Consciousness represents an innate framework in which complex, flexible and adaptable biological decision systems are `allowed' or `forced' to adaptive self-government.* The flow of experience is the flow of information that befalls a particular survival machine and that is `weighed' according to inborn values which represent the interests and perspective of its genes. Consciousness is the way in which information is thus `subjectivized' to make autonomy possible, within limits that guarantee that the organism remains dedicated to its genetic mission.

Pugh's model thus revolutionizes our way of understanding consciousness. According to Pugh `the cognitive decision process is intimately linked with our sense of awareness or consciousness' (Pugh, 1978: 154). Consciousness can be termed the way in which we are *forced* to *use* information to make decisions. Consciousness is the product of the `experiential values' which are programmed into the individual organism as a result of a long evolutionary history and which are orchestrated together in such a way that it is able to make decisions which are in its own interest or in that of its genes.

3.4 CAPTURED WITHIN OUR COCKPITS, NAILED ONTO OUR DASHBOARDS: THE RELATION BETWEEN BODY AND SELF

Despite the fact that they give, by definition, incomplete knowledge, metaphors can be very instructive. They often show analogies across different realms of knowledge and enable one to give names to phenomena about which it would be otherwise very hard to talk. Of course, it is important to be aware of the limits of the metaphors one uses: One should leave one's ship the moment it grounds.

With respect to consciousness, we now have an interesting set of metaphors. If we combine Dawkins' metaphor of the survival machine and Pugh's analyses of value-driven decision systems a whole set of interesting new metaphors about the mind emerges (Slurink, 1986). The organism can be seen as an organic missile, a smart bomb, that did not fire itself, but has been globally instructed with a set of selected values with which it takes its `own' course in a new environment. It does not know that the way in which it takes this `own' course is in fact, via these values, based on a long history of success and failure. Somewhere in the center of the missile, say in its cockpit, a decision center emerges from which `it' views the world and plans its course. Inside this cockpit a simplified map of the world helps `it' to orient itself, a series of bulbs, switches and monitors shows `it' the dangers, hopes and possibilities.

Where, then, is the person *behind* this organic dashboard of bulbs, monitors and switches? Who is it and can it leave its cockpit? No, if this were the case many survival machines would probably be deserted in the midst of the struggle for life and they would never accomplish their genetic mission. Even heart-attacks and suicides cannot be interpreted as desertion, but should be seen as accidents or the turning of a switch that exists only for emergencies. There is no metaphor for the relation of an organic dashboard and its pilot, because our artificial robots are still too primitive. The pilot or driver inside an organic robot, the `I' behind its dashboard, is in fact *this dashboard itself which is curved and monitors its own monitor.* In vain are we looking for an independent homunculus inside the cockpit: The only unique `homunculus' inside is *a particular point of view, the source of a* *unique stream of experiences, another version of the world-movie as it can only be seen from somewhere `within'.* Physically this is probably an emergent property of the `recurrent networks' and `loops' between thalamus and cortex as noted by the connectionists and loop theorists.

What about self-consciousness? Should this not be seen as an extra dimension resulting from some sort of extra ability? Yes; as a result of our evolutionary history as highly social animals in very complicated societies we, as humans, develop a self-image during our lifetime. This self-image enables us to guess how our conspecifics judge us and is therefore often strongly linked to evaluations of merit and rank. However, even this self-image should be seen as another lifelike phantom projected on the `monitor' at the inside of the organic cockpit which is the locus of our subjectivity. Maybe we should speak about a special extra monitor on which social relations are mapped and which gradually starts to represent its own host as a unique separate individual, an `I' in a world of `you's'. Anyway, we do not see ourselves from the point of view of an extraterrestrial, `objective' ethologist most of the time; we simply play a role in most of the movies on our own monitors. (Of course, this opens up the possibility of an infinite regression of reflective loops, but our neurocomputational systems do not break down as a result of such loops, they simply get bored and start to pay attention to other monitors).

Does this view of consciousness help us to resolve the classical philosophical topics relating to that subject? A philosophical question deserves a philosophical answer: yes and no. Yes:

(a) - It helps us to resolve the mind-body problem: The *cause* of the classical dualism is simply the reification of two points of view relating to the same body: the unique private view from `within' (the perspective from which a survival machine calculates its interests, the internal `dashboard' of subjectivity) and the much more public view from the outside. (Exit both simplistic dualism ànd monism.)

(b) - It helps us to explain why there is a gap between the world of experiences - *die Erscheinung* - and the unknowable world `itself' - *das Ding an sich*: Information relating to our world has to be `subjectivized', because only in this way does it help us to make adaptive decisions. (Exit naive realism and idealism.)

(c) - It helps us to explain why there is an unparalleled type of causality, agential causality (R.W. Sellars, 1973), which seems to be directed `downward' (`downward causation'; Campbell, 1974). This `top-down' type of causality is to be expected where a number of different input channels have to be integrated in some central *locus* of weighing, scenario-building and deciding. Of course, this `top-down' causality is intimately linked with a whole battery of `bottom-up' channels via which the input from the external world is gradually selected, interpreted and valued. Finally all this information arrives at the central `dashboard' to be weighed together with other generalized information to allow balanced decisions. Thus, bottom-up and top-down causality *together* enable the organism to cope with its environment and to remain faithful to its genetic mission at the same time. (Epiphenomenalism and the QQ-thesis are therefore implausible; exit both hard determinism and the indeterministic theory of free will.)

(d) - It shows us the origin and nature of values, which might ultimately be called `subjective', but which we share with our own conspecifics and which we also share, to some extent, with a series of other species and to which we therefore can refer as though they belong to an `objective' world. They do not have their origin in a platonic sphere beyond actual organisms, nor do they have their origin in some rational contract between them. Nor do more `sophisticated' values, like curiosity or a sense of beauty, necessarily reduce to more `primitive' values, like pleasure and pain. The fact that values vary throughout different cultures and times (a subject to which we return in chapter 6) does not exclude the possibility of the existence of an innate human value structure which is, roughly speaking, universal to members of anatomically modern humans. Ethical relativism and voluntarism have to take into consideration both the effects of these semi-universal values, which derive from the innate `structures of experience', and the way in which they lead to a public `morality' as a result of particular ecological pressures and social `system requirements'. (Exit both moral objectivism and a too-simplistic relativism.)

(e) - It may help us to assess the question of animal awareness more realistically. Pugh's `value-driven decision system' view suggests a gradualism with respect to the level of consciousness displayed by various animals and by humans of different ages and talents. Just as there exist screw-drivers of all types and sizes, there probably exist different classes of intelligence which may fit into different structures of the world. As is to be expected with respect to a great variety of different vehicles with different purposes moving around in different terrains, each vehicle has its own type of dashboard which affords it an incomparable window on the world. (Exit anthropocentrism; more on this question in 3.6 and 3.7.)

However, we should also pay attention to the `no' answer. No:

(a) - Of course, we can only answer such classical philosophical questions *by changing them slightly*, by consciously neglecting and transcending their implicit ontologies. Therefore the traditional philosopher can always retort that the naturalistic answer does not match his question.

(b) - At this point the answer to the `why'-question relating to consciousness is only schematic and largely intuitive. Perhaps it is true that consciousness evolved *to force organic vehicles into making semi-autonomous adaptive decisions*, but then, still, we do not know why pain, pleasure and related sensitivities were the only evolutionary option for such a compulsion and how neural networks have to be wired to get such results.

(c) - Finally, it is quite possible that our natural categories are simply inadequate to understand complex emergent properties. Our intelligence, which seems partly to have co-evolved with our technological skills, is often better at analytical reduction (bottom-up analysis) than in understanding a system as a whole and the way in which particular properties emerge as a result of its specific composition.

A typical expression of the tendency to prefer analytical reduction to synthetic reconstruction is Churchland's claim that we can and should learn to describe our qualitative experiences in terms of neurophysiology. Our experiences are constituted both by the unique objective situations in which we find ourselves and by our unique subjective evaluations of them. Of course, they are embodied in a transient state of our neural network as it interacts with the world via a battery of sense organs. It is uncertain, however, whether it would be possible to describe them sufficiently by just isolating the neural network from the rest of the world and noting the configuration of its synaptic weights at one particular moment. The essence of those experiences probably lies in the dynamic brain-world interaction of which they form a part and *in the way this interaction is represented `from within' our `phenomenological cockpit'.* Introspection can never give us our qualitative experiences as they `really are' in neurophysiological terms, but only as they `really are' on our `internal monitor', as it is designed by evolution to control our cognitive and evaluative interactions with the world. Qualia are not simply neurophysiological states, but the way in which we experience and evaluate a particular situation `from within' an ongoing series of those states. There is no need to `translate' them in neurophysiological terms, because we can already feel what they really are. One cannot describe a movie in terms of the workings of a television, although the television certainly is a factor in what a movie looks like; and, of course, we are not only watching a movie, but playing a role in one as well.

Consciousness as the autoconnected dashboard of the mind. In all, there are good reasons to interpret consciousness as an emergent property, the product of variation and selection of `holistic' properties of *whole* organisms, embodiments of a history of successful behavior in a series of past environments. Its function or evolutionary *raison d'être* is probably that it enables organisms to cope with unique, unprecedented situations; as a result conscious states are probably as various as the brains of actual organisms and the situations in which they find themselves. I have proposed, in the spirit of George E. Pugh, that consciousness represents the way in which information is `weighed' in function of its survival value and have used the metaphor of a dashboard to show how it is linked to decision making and how it gives direction to the interaction of a particular survival machine and its environment *via* a simplified representation within a rigid, innate frame-work. This metaphor shows at the same time that consciousness is not a kind of accidental intrapsychological transparency, but that it is *designed* in a functional way to make adaptive decisions to new and unique situations possible.

Some characteristics of consciousness can thus be explained:

- *Qualia* and qualitative distinctions arise, because sophisticated decisions in complex environments require a differentiated set of sometimes opposed values (as artificial decision systems show). The emotions seem to be orchestrated such that information *has to* reverberate inside the organism into adaptive decisions. Subjectivity thus constitutes a kind of sensitive interface between the genetic interests of a particular survival machine and its environment such that this survival machine is *forced* to take particular types of information seriously. From Pugh we learned that a value-driven decision system that has to be programmed to perform certain tasks may not be allowed to change its own `primary' values.
- *Intentionality* is an effect of the preprogrammed goal-directedness of an organic decision system as it guides the interaction with the environment from `within', from behind its `dashboard'. It constitutes the form in which the representations on the various monitors and dials *refer* to the real world outside and to the virtual targets and goals of the organic vehicle. The properties of the environment that are used for orientation have to be selected and valued to create a schematized mapping inside the neural network that enables it to build scenarios and to choose between them. Because the organism often needs specific information its sensors need the ability to focus on the outside world which is only possible via a feedback loop in which the representations `on the inside' guide the systems of information-gathering in the outside world. This is only possible, of course, if these representations link up somehow with the real world. The mysterious `aboutness' over which philosophers have pondered for several millennia, is a characteristic of this process of focussing, in which internal representations are gradually improved by manipulating the

external sensors and feelers.

- *Control* in unprecedented circumstances is the primary function of consciousness. Control implies a cognitive assessment of a completely new situation in which one often has to take unprecedented steps to remain faithful to one's goals. As Pugh has shown a complex set of heuristic values can be an ideal device for estimating the consequences of various possible decisions in such a situation. This only confirms the idea that control and `freedom of the will' is not opposed to causal determination, but to the dogmatic activation of rigidly preprogrammed action patterns and to the inability to devise entirely novel solutions, plans and decisions. (Freedom is therefore a relative notion, entirely linked to one's elbow-room before one's drives).
- Consciousness has the character of a *stream*, because an organism in its environment has to deal with a lot of independent problems, some more urgent than others, and at the same time still needs all its cybernetic resources in order to devise an optimal solution to the most important problems. Apparently this `concentration' of effort is needed to pre-activate all relevant prototype vectors and knowledge and to ensure that all relevant values are weighed in one shot (perhaps a shot in which a whole web of neurons begins oscillating in a 40 Hertz rhythm). Although all organisms have many problems, creative solutions require an answer which arises from the organism as a whole. Less urgent decisions can be postponed or delegated to subconscious processing. Often consciousness is only needed to initiate search procedures, so the stream-character of consciousness is not in opposition to a certain level of multi-tasking. In nature, an organism has to be continually watchful of any new dangers and possibilities that it may encounter. A whole battery of sensory canals has to be kept open all the time and urgent information from the external world should be able to overrule all current conscious activities. In that respect consciousness is similar to the changing field of attention of the pilot behind a dashboard, who will sometimes turn to his attention to one monitor, then to another, then to a flickering bulb, etc. The autoconnected dashboard is a place where different problems are continually fighting for attention and for the cybernetic resources that go with a centralized decision procedure.
- *Self-consciousness* could be a more relative notion than is often thought, because the `self' which is represented on the internal monitor can be viewed from several viewpoints. A cat which is continually cleaning itself needs a representation of its own body contours and fur; a cat that has to decide whether to fight with the neighbor's cat has to make an estimation of its own strength and condition. A macaque that approaches a water hole, at which some group members are already quenching their thirst, needs to make a complex calculation in which its own sex, age and rank are compared to that of the other animals that are present. Humphrey, Trivers and Alexander, amongst others, have speculated that a sophisticated self-consciousness is the result of adaptations that enabled our ancestors to cope with extremely complex social environments. It should probably be seen as the result of genetic instructions which guide the growth of the brain in such a way that cybernetic resources can be spent on the representation of complex social networks. At some point (perhaps in the Pongidae and Hominidae only), not only is there a representation of the self and of others, but at the same time a representation of the self *as viewed by others.* This social self-image can then be manipulated continually not only by others, but also by the self (thus the effectiveness of `positive thinking',

self-hypnosis and some forms of psychotherapy).

3.5 THE ADEQUACY OF THE DASHBOARD METAPHOR: SOME LINKS TO THE EVIDENCE

Recent new discoveries have strengthened the idea of an innate dashboard on which information from different canals and brain-regions is centralized to allow for optimal decision-making. In a discussion on phantom limbs, Ronald Melzack reports that even people born without a limb often perceive one from time to time. He tells us about an eight-year-old boy, born with paralyzed legs and a right arm that ends at the elbow, who sometimes feels phantom fingers. Another example is a 32-year-old engineer, born without a leg below the knee, who often experiences a whole leg, including a foot. Sometimes the experience will disappear only to return, to his relief. Melzack proposes that the brain contains a neuromatrix, or network of neurons, that, in addition to responding to sensory stimulation, continuously generates a characteristic pattern of impulses indicating that the body is intact and unequivocally one's own, the `neurosignature'. If the neuromatrix operates in the absence of sensory inputs from the periphery of the body, it may create the impression of having a limb even after that limb has been removed. (Incidentally, temporary states of this neuromatrix could also explain the sensation of rising outside of the body and returning into it, P.S.).

Because phantom limbs are sometimes experienced by people who never had the original limbs or who lost them at an early age, Melzack proposes that the neuromatrix is largely genetically prewired, although it can also be sculpted by experience. He believes that the brain produces a natural `body image' which does not stop working the moment that external input stops (Melzack, 1992). Of course, the brain produces more than a body image alone. It also contains a natural `world image' which is normally constructed and corrected on the basis of sensory input. If input from the senses stops coming, however, the brain may simply create an image itself. This seems to be what happens when we dream, but this could also explain the phenomenon of *phantom seeing and hearing* about which Melzack also has something to tell us. Phantom seeing and hearing often occurs in partially blind or deaf people. A lack of input from eyes and ears causes the brain to generate very vivid images and sounds itself. One woman who is partially blind continues to see a phantom building, which will come and go unexpectedly. Another woman who was a musician before losing her hearing hears piano concertos and sonatas which she cannot turn off and which sound so real that she first thought them to come from a neighbor's radio.

Both the body image and the world image can be seen as projections on `internal screens' which have been erected by evolution to give us concise information on the basis of which we can make decisions. Several authors have proposed similar metaphors to account for consciousness and sometimes they even try to locate the transition to consciousness. With respect to vision, we have seen already in chapter 2 that Erich Harth points to the LGN as an `internal sketchpad' (Harth, 1995: 70). With respect to emotions and to qualia, Pugh tries to explain why they are qualitative rather than quantitative by pointing to the relative distinguishability of qualitative signals. According to him, the use of distinguishable different values makes it easier to associate specific value components with specific causal factors (Pugh, 1978: 110). Implicitly he is comparing the framework of our consciousness here with a kind of dashboard on which the different monitors, bulbs and measuring-instruments have to be conveniently arranged and clearly indicated to make decisions possible.

The dashboard metaphor is also compatible with the distinction between conscious and unconscious. There is a lot of activity in the organism that has to go on unconsciously and in which we have no say. If we could stop our own heart-beat, many would stop it if they found themselves in pitiful circumstances. If we could reprogram ourselves so that we would be happy under any circumstances in which we may find ourselves (with or without a particular partner, with or without a job, in absolute poverty, after a tragic loss, etc.), we would probably reprogram ourselves in such a way that we would not be able to feel pain and sadness anymore and that we would feel joy, satisfaction and ecstasy even in the midst of disaster - as a result of which we would probably become extinct within one generation. Therefore, most of the machinery behind our internal monitors and bulbs is completely sealed away from us and we can only fumble around with some of our wheels and switches. In that respect we are in a deep sense captured within our cockpits and nailed onto our dashboards. Yes, we are free, we make choices continually, but at the same time: no, our freedom is limited to the options as they are given emotional values by our partly innate value-system and as they are presented to us on our `internal monitors'. The degree to which we are free is largely dictated by the adjustability of our monitors and the options on our switchboard. We cannot change the way in which we are conscious about the world.

Finally, psychopathology brings out the strength of the dashboard metaphor. Patients with different psychological abnormalities can be viewed as persons without specific monitors or with monitors that give a distorted view of reality or that represent aspects of the world to which a normal person does not pay much attention. Autism is currently understood as an impairment of a specialized, largely innate module, situated in the left medial prefrontal cortex, which enables us to reconstruct mental states of other persons (e.g. Frith, 1993). Depression can be viewed as a state in which an overactive amygdala prompts the left prefrontal cortex to devise dark scenarios and gloomy or one-sided views about other people (Posner & Raichle, 1994). The dashboard metaphor shows very clearly that people suffering from such illnesses live in their own idiosyncratic worlds, resulting from lack of particular monitors, or from their oversized format, while at the same time stressing that the properties of such worlds can be studied and explained scientifically.

3.6 CONSCIOUSNESS AND ITS ANATOMICAL AND BEHAVIORAL CORRELATES

In spite of the fact that consciousness is a private property, it does not necessarily have to remain completely mysterious. I have suggested that it was designed to force animals to take into account multiple factors in flexible decisions. Qualia are necessary to reward and punish behavioral experiments in such a way that an organism can gradually acquire a series of behavioral patterns which are adequate both to its needs and to its environment. Play behavior may be a sign that this type of learning is present in a species. Intentionality is necessary as long as an organism has to be guided by temporary plans and goals, which have to be adjusted each time depending on its situation and its needs. Intentionality may be interpreted as a characteristic of an active mind, which does not simply react to stimuli, but that fulfills its mission via a series of temporary plans which are the product of its imagination. The ability to conceive such plans and to `write' such scenarios may manifest itself in idiosyncratic behavior, but it could also manifest itself in the ability to dream, dreaming being conceived here as the activity of a mind which is continually scanning possible scenarios in a situation in which the information from its sense-organs is not passed on.

Consciousness, therefore, according to this model, is linked to choice and values. As our `innate structures of experience' are the product of variation and selection of a long series of past choices they can be considered as sediments of past experiences and past lives. Ancestral animals which tended to make adequate choices as a result of adequate sentiments simply contributed more to the design of the psyches of their modern descendants. Evolutionary `monadology' can be seen therefore, paradoxically, as a doctrine about selective `metempsychosis'!

One of the problems of the theory of metempsychosis was, however, that it remained unknown whether animals could reincarnate into humans or not. In later antiquity an increasing number of philosophers started to claim that only rational creatures could reincarnate into each other. Yet, consciousness could be something more fundamental than rationality, which can be claimed to depend at last partially on language. Perhaps all animals exhibit some degree of consciousness. At least the word animal derives from *anima* or soul. Do bumblebees, cuttle-fish, eels, salamanders, seven-months-old embryos, penguins, elephants and dolphins *experience* their lives to some extent consciously?

The problem of animal awareness is one of the hardest tests for each model of consciousness. The problem with animals is that they do not speak, or, at least, are unable to relate to us their inner experiences. Thus, if we want to answer the question of whether they are conscious and what kind of consciousness they do possess we have to look for the anatomical and behavioral correlates of specific types of consciousness. Can the above model help us find these?

Qualia. In the spirit of Pugh qualia constitute the heuristic values in the organic decision system. Why do they have to be *experienced*? Perhaps the simple reason is that they would otherwise be neglected. *The intensity of experience therefore has to correlate to the urgency of the decisions that have to be made (in terms of selfish-gene interests) insofar as this urgency can be somehow assessed by the organism in question.*

What animals *do* experience them and to what degree? *As qualitative experiences seem to be designed to enable the organism to `measure' the adaptive value of a large number of behavioral alternatives, they are probably experienced by animals to the extent in which they display flexibility and creativity in their natural environment.* Of course, it is not easy to assess the degree of this flexibility and creativity in the laboratory. Because the `primary' values have to be innate, it is to be expected that individuals that are unable to display their `creative potential' as a result of special circumstances are not falling short of the `experiential level' of their species. As the `primary' values are a product of evolution, we have to expect that they are at least as completely differently orchestrated in different species as they are already differently orchestrated in different individuals of our own species. Species that, even in their natural environment, do display a lot of rigid preprogrammed action patterns, which cannot be corrected by taking into account environmental novelties (a wasp that does not think about turning away if hitting a glass panel), probably do not use subjective experiences in the `decisions' that constitute those fixed action patterns. Species which have evolved to be more flexible, however, need not to have lost all of their more rigidly preprogrammed action patterns. Instead, we would expect a kind of hierarchy of psychological mechanisms: a., in which some reflexes guarantee a minimum security level, b., in which a series of preprogrammed stereotypical action patterns constitute the behavioral repertory which is typical for the species (which however in many cases has to be refined by training) and c., in which, in only some specific domains, more flexibility can be built in by learning which results in completely new patterns of behavior.

These learning processes probably enclose conscious experiences with which the trials are rewarded and the errors punished. Of course, between the level of fixed action patterns and learning processes there is also a level of acquired habits which result from past learning processes and which allows for their semi-automatic execution. *Apparently the procedure of conscious `weighing' integrated information and behavioral scenarios is followed only if fixed action patterns and habits are inadequate for the correct performance of the survival tasks of the organism*. Often the results of such conscious learning processes are `stored' as semi-automatic action patterns or habits.

Thus, we would expect conscious experience especially in animals which display a lot of creativity in their natural environment and which are able to change their action patterns at the moment that these become unadaptive. Let us not forget, however, that even humans display a lot of unadaptive behavior, especially if this is reinforced by traditions which have been formed in other ecological circumstances. The hall-mark of consciousness, however, seems to be constituted by learning processes on the basis of the creative testing of new behavioral possibilities. *It is probable that species which display play-behavior, at least in their juvenile phase, are conscious.* Such behavior is demonstrated by many birds and mammals.

There are several reasons to expect consciousness especially in animals with parental care. One of them is purely economical: in these species individuals are more `valuable' (they are more `K selected') and any extra education would be a useful supplement to their expensive upbringing, increasing the likelihood of their success. Another reason is simply practical: in these species parents and offspring meet and therefore, at least the possibility exists of transmitting valuable non-hereditary information from one generation to another. As long as the parents are able to create an environment of relative safety, the juveniles of such species can practice a number of skills under the supervision of their parents. For example, in many species of birds and mammals hunting skills improve only gradually while parents gradually reduce their level of feeding and protection over an extended period (consider Trivers' `parent-offspring conflict', Trivers, 1985, chapter 4.3). In these species a lot of behavioral experimentation can take place during this period of extended dependence. It is probably in these species in which we should expect to see the crucial role of play behavior in the gradual genesis of the adult behavioral repertory and in which we, therefore, should expect to see the crucial role of consciousness, too.

Intentionality. Apparently intentionality concerns the way in which objects or relations surrounding an organism are `represented' in its mind and in which it can manipulate both these objects or relations and itself to improve these representations. One can speculate whether this kind of `intentionality' presupposes consciousness. Many modern cameras are equipped with an auto-focus system which via a feedback loop guarantees that the object is projected sharply on the film. Of course, this kind of focussing can work entirely without consciousness.

Probably something else is meant by Brentano c.s.. Intentionality seems to refer to an orientation to goals: not to goals which are somehow fixed, but goals that can be fixed entirely anew at any moment. Intentionality seems to be an `open' goal-directedness. This is entirely compatible with the element of scenario-building in the value-driven decision system. Intentionality seems to be specifically attributable to organisms that are continually changing their internal goals and `weighing' the value of the goals they are imagining and striving for.

How could we know which animals are such natural `scenario-builders'? One possibility is that we can simply observe the animal in question and see how it seems to change its goals in the middle of its activities. This procedure can be hazardous, however, because this change of behavior could also result from a clash of two fixed action patterns. Another possibility is that *dreaming* can be interpreted as the activity in which free course is given to the mind without feedback from the environment. Animals that dream are animals with a mind that is continually throwing up and going through possible scenarios as a result of which they can better anticipate changes in their environment. Whether an animal is dreaming or not can often be observed as a result of grunts and rapid eye movements while it is asleep (REM sleep). According to Winson (1990) placentals and marsupials do exhibit REM sleep, whereas more `primitive' mammals, including the echidna, do not.

Volition. Many animals do experience conflicting impulses, for example when a male is unable to decide whether to fight or court a female. Often such animals will display so-called displacement activities which to an outside observer seem totally irrelevant. They will start scratching themselves, or start pecking into the ground, etc. (Humans are believed to show displacement activities, too, for example in the form of sucking on pens and spectacles, head scratching, beard stroking, McFarland, 1987.) In many species of birds such displacement activities have evolved into ritualized behavior.

Conscious control over one's own behavior perhaps presupposes a behavioral flexibility which enables one to solve conflicting impulses in a more creative fashion. This can only evolve the moment an animal is able to postpone the execution of some of its drives or the moment it is able to transform or integrate its drives. True control is only achieved if the animal is able to give priority to some goals over others or if it is able to devise new goals and plans in order to integrate some mutually almost incompatible drives. It presupposes the ability to devise internal models of the possible results of one's own different possible strategies. It is known that chimpanzees, at least, are able to suppress short-term desires in favor of the achievement of their goals. Frans de Waal tell us, for example, how one chimpanzee pretended not to know where some bananas were buried only to unearth them the moment no other colony members were watching (de Waal, 1981).

Such conscious control over one's actions probably transforms instinctive drives into a system of values which can be used to weigh alternative action-plans. Even human beings, however, do not always think before they act. `Free will' is only a relative notion: the stronger the ability to predict the consequences of one's own actions and to adjust and change one's plans creatively, the freer one is. It seems to me that most of us are not always as free as we would like to be.

Consciousness should not be thought to be linked to a free will, however, but to the experience of qualia. An animal that experiences qualia is able to learn from its experiences, but it may also have to endure experiences from which nothing more can be learned than that the world is not always pleasant and just.

Self-consciousness. In one form or another consciousness can always be viewed as self-consciousness, because it is always one particular point of view that is experienced and many animals need to know the effects of movements of their own bodies. A squirrel has to guess its own weight, a zebra has to know when her young is standing in her own shadow, etc. Additionally, all social animals which live in a hierarchy that is not fixed, but results from competition, seem to need some kind of image of their own social status. Individuals of such species which are consistently defeated in encounters with their opponents often become psychologically `down', display timidity in encounters with new opponents and will, as a result, remain in their position of low rank. This effect can even be observed in bumblebees and crickets (Wilson, 1980: 123), although, of course, it is unclear what kind of self-image is implied. Of course, one possible test that an individual is aware of its own status would be the purposeful use of signals and ornamentation to advertise one's status. Recently, it has been discovered that the reddish taint of the breast-feathers of the Lammergeier is not only placed there by themselves, but is also a clear signal of its status.

At the moment many authors seem to agree that self-consciousness, as it is expressed in apes and humans, is probably a result of social evolution. The idea is that a self-image enables us to guess how we are judged by conspecifics and, therefore, also enables us to manipulate them better. A good argument for this position is that our self-image always seems to reflect social value-judgments and that the way in which we relate to ourselves, even in diaries and prayers, is tightly linked to judgments concerning `merit' and `self-esteem'. A closely related argument is afforded by the theory of reciprocal altruism: it is predicted that it is adaptive to deceive oneself and others about one's investments in the reciprocal relationship. To be able to do this one needs on the one hand a flattering self-image - on the other hand one needs to `know', unconsciously, one's real purposes (Trivers, 1985; 1991; `knowing unconsciously', of course, is an interesting paradox). The ability to recognize oneself in mirrors, which can be learned by chimpanzees, orangutans, and dolphins, but not by most gorillas and not by any monkey (Parker *et al.*, 1994), could be a side-effect, perhaps also related to a talent for object manipulation and rotation.

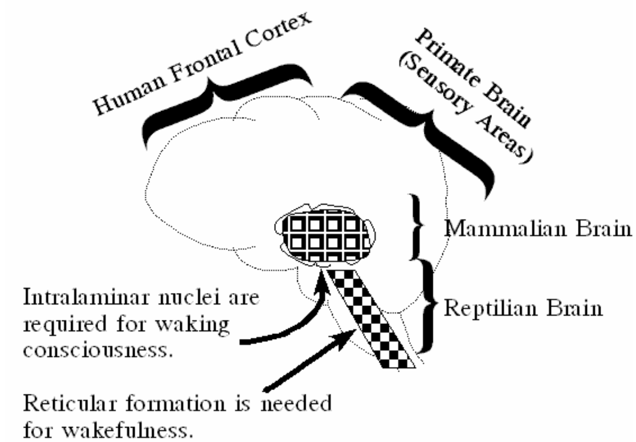

Figure 3-4. To the extent that animals share particular parts of our brains known to have specific functions, it is reasonable to believe that they share the corresponding experiences. Redrawn after Baars, 1997: 32.

Anatomical evidence. Often anatomical evidence offers a invaluable amplification of the behavioral evidence (the argumentation used is a kind of triangulation). For example, both corvids and parrots are known to be groups of birds displaying an extreme level of playfulness and intelligence. (I recently watched a crow trying to open an acorn by dropping it repeatedly on the road). At the same time, it is known that both groups of birds have a relative big *hyperstriatum*, which is the part of the forebrain responsible for general intelligence in birds (Savage, 1995). Above that, corvids are at the top among birds with respect to brain size, having a brain-to-body ratio equalling that of dolphins.

If we use this kind of anatomical evidence and study the evolution of the brain in our own lineage, there is are good reasons to believe that we share the division between waking and sleeping with reptiles, that we share the behavioral flexibility which comes with an elaborate system of values and emotions with all mammals and that we share our detailed visual world with all primates (Baars, 1997; fig. 3-3).

Conclusion. In what kind of animals may we expect consciousness and the ability to suffer? It seems that the ability to use symbols and language and to recognize one-self in mirrors are neither necessary nor sufficient for the ability to suffer. The ability to suffer seems directly implicated by the ability to experience qualia. (Qualia seem to be linked to the type of behavioral flexibility we see at least in birds and mammals.) More sophisticated psychological abilities could enhance it, but they could also give some relief.