Andreas Blank Editor

Animals

New Essays

Philosophia

Basic Philosophical Concepts

Consciousness as an adaptation - what animals feel & why[1]

Pouwel Slurink

> "While the behaviorists recognized that aversive UCS (Unconditioned Stimulus), such as predator odor or electric shock, can serve as 'punishments' in many learning tasks, they could overlook as irrelevant that fact that UCRs, such as fearfulness, also have an internal feel to them. Other UCS could serve as 'rewards' that would promote the learning of approach behaviors rather than avoidance behaviors. There has traditionally been little discussion, however, on any corresponding feelings underlying the logic of behavioral learning in animals. Of course, it is likely that rewards and punishments only work so well to control learning because they generate affective feelings in the brain. The spooky process of *reinforcement* may reflect the way feelings work in the brain."
> J. Panksepp & L. Biven, *The Archeology of Mind,* 2012: 12.

Summary

Which animals have consciousness and which do not? To answer this question we first need a theory about the nature of consciousness. It will probably always remain somewhat of a miracle how consciousness arises from the synchronized firing of groups of neurons. Nonetheless, we currently have sufficient knowledge of brains and evolution to conclude that consciousness is a perfectly natural phenomenon. Additionally, from an evolutionary epistemological and behavioral ecological perspective it is relatively easy to understand the *function* of consciousness and why there are different 'kinds of minds'. In this paper it is argued that consciousness arose as an evaluative framework which enables mobile organisms to make adaptive decisions. The way we experience the world is already a product of evolution. This evolutionary approach frees us from dualism and anthropocentrism and sheds new light on the relationship between consciousness, reason, the emotions, and choice. It is consistent with both Dehaene's 'neuronal workspace theory' and Panksepp's 'affective neuroscience'. Above that, it offers a framework from which we can deduce which types of animals are almost certainly able to experience different kinds of sensations and which types of animals do probably not.

Introduction

Generally speaking, most western philosophers have had considerable difficulty in conceptualizing different 'kinds of minds' (Dennett, 1996). The main cause seems to be a preoccupation with our own special position and responsibility and the dominance of a dualistic, religious Platonic-Christian interpretation thereof. In Platonism and Hinduism a dualistic belief in reincarnation does not exclude a gradualism with respect to the relation between man and the other animals. In early Christianity, however, the uniqueness of each individual human soul is stressed at the expense of the experience of continuity and connectedness with other living beings. The notions of an exclusive relationship between man and (a transcendent) God, and of a largely divine but in part human mission to change

[1] In: Andreas Blank (editor), *Animals, New Essays*. Philosophia Verlag, Munich, 2016, pp.303-327.

the world, tipped the balance strongly in favor of anthropocentrism. Of course, one of the advantages of this position was that it enabled their proponents to defend (their) morality as a God-given, absolute phenomenon.

The transition from ancient to Christian thinking about animals is beautifully documented in Origen's *Contra Celsum* (circa 248 a.d.; Chadwick, 1980). As a typical ancient thinker, Celsus attributes an enormous memory to elephants, linguistic talents to ants, and foresight to birds. In contrast, Origen makes an absolute distinction between animals and 'rational souls'. Only 'rational souls' are sufficiently disembodied to be able to move from body to body. In his *Peri Archon* (circa 212-215 a.d.; Sieben, 1977) he presents the material world as a kind of dungeon in which fallen angels and other rational souls can find their way back to righteousness. That implies, at the same time, that the world was created exclusively *for* those rational souls and their bliss.

Of course, Origen was considered a heretic and not all western philosophers were as extreme in their opinions on the differences between animals and man as he was. Yet, the identification of mind-body dualism with a dualism between man and other animals is repeated time and again in Western philosophy. The same goes for the notion that man is a unique being with the unique ability to transcend his own impulses and even all natural causality – think especially of Descartes and Kant. Even the 'evolutionary' philosopher Bergson, with his *'Évolution créatrice'* (Bergson, 1908), still tries to find holes in the chains of necessity and to unite evolution, spirituality, human uniqueness, free will, and moral progress ('Dieu se fait').

Darwin put us back with two (not four) feet on the ground. His 'theory of descent with modification through natural selection' enables us to find a more balanced view of the relationship between man and other species. In contrast to Lamarck's theory of evolution, Darwin's theory does not imply mysterious 'nerve energies' and a disguised notion of progress. After his trip around the world on the Beagle, the young Darwin became convinced of the 'transmutation of species' and decided to search for a completely naturalistic – almost geological - mechanism behind it. After several years, he "happened to read for amusement" Malthus' treatise on population (in July, 1838; Darwin, 1958). He immediately realized that the notion of a 'struggle for life' resulting from overpopulation was exactly what he had been looking for all the time. He realized that, when this struggle goes on in all life and is combined with tiny hereditary differences among individuals, this would result in the survival of those varieties which are slightly better adapted to the current circumstances than others. This simple combination of reproduction, variation and selection would then be enough to explain 'the transmutation of species'.

Why is this insight so important? Firstly, it gives us a concise explanation of evolution that links biology, and even psychology, to the rest of the natural sciences. Secondly, it indicates that many properties of both animals and humans are either adaptations or result from them. New features, including those that constitute human uniqueness, do not come out of the blue, but arise from necessity: mutants with them somehow fare better than individuals without them. Thus, if man is unique – as is each species – his unique properties

must have originally constituted an advantage above other properties in our ancestral environments ('the environment of evolutionary adaptedness', according to Bowlby, 1980).

Thus, instead of simply stating that man is 'just another ape', Darwin's theory implies that we can rephrase discussions about human uniqueness and about the differences between man and other animals in terms of adaptations that are either shared - as a result of common evolutionary pathways - or are uniquely human - as a result of a uniquely human evolutionary trajectory. This way, a much more nuanced view about the talents of different kinds of animals is possible as well as a deeper understanding of human uniqueness. Of course, this kind of understanding requires an enormous knowledge about the conditions in which apes, hominids and humans evolved and to which they are adapted to some extent. Although much progress has been made in this area, a series of mysteries is likely to remain here as well.

Reductionist naturalism: from dualism-anxiety to naive objectivism

Of course, the Darwinian theory of evolution has been around for a while and there have been many attempts at deriving models of animal and human cognition from it. During the twentieth century, however, both science and philosophy were still often too much dominated by opposing metaphysical views to be able to appreciate or even understand the advantages of a Darwinian approach. On the one side, positivists and physicalistic naturalists tended to view the very concept of 'consciousness' as a remnant of dualistic metaphysics – hence the popularity of behaviorism with its dogmatic rejection of both empathy and introspection as sources of knowledge. On the other side, phenomenological and hermeneutical philosophers tended to dismiss naturalism altogether in the name of pure 'experience' or, at a later stage, some version of sophisticated relativism. Only a small collection of independent thinkers, like Roy Wood Sellars and Konrad Lorenz, managed to keep a cool head during these heated debates and managed to transcend the presuppositions leading to its false oppositions. During the computer age, behaviorism was finally succeeded by cognitivism, but this again led to a one-sided model of the mind as a purely passive 'information processing' organ.

As a result, even philosophers that think of themselves as naturalists and Darwinists are sometimes still producing models of cognitive evolution that do not address the question of functionality adequately. Even 'naturalists' like Dennett and Churchland suffer from a dualism-anxiety which evokes memories of the logical positivists. Let me illustrate this with Dennett's model of the 'various design options for brains', which is in fact a kind of rational reconstruction of cognitive evolution. Somewhat pretentious and bombastic Dennett speaks of 'the Tower of Generate-and-Test' (Dennett, 1996)

According to Dennett, there are about four floors in this cognitive tower and each new floor reflects a new level of behavioral sophistication. "As each new floor of the Tower gets constructed, it empowers the organisms at that level to find better and better moves, and find them more efficiently". At the ground floor we find *Darwinian creatures,* which are

completely ‘field-tested’ and unable to correct their own behaviour during their short lives. Any small error in their innate behavioral programs may lead to their premature death.

The next floor is inhabited by creatures with built-in ‘reinforcers’ that enable them to learn ‘smart moves’ by a “long, steady process of training and shaping by the environment’’. “These individuals … confronted… the environment by generating a variety of actions, which they tried out, one by one, until they found one that worked”. Dennett calls them *Skinnerian creatures* and claims that Hume’s associationism was a predecessor of both behaviorism and connectionism.

The next floor is reserved for what Dennett calls *Popperian creatures,* which have some kind of internal maps or representation of the world in which they can try out their behavioral experiments virtually. Strikingly, but perhaps characteristically, Dennett evades notions of ‘innate ideas’ in this context, or of the ‘a priori’. The only thing he deems necessary are “perceptual mechanisms designed to ignore most of the flux of stimuli”, which “concentrate on the most useful, most reliable information”.

Dennett’s last floor is reserved for organisms that learn from one another. Dennett calls them *Gregorian creatures,* in honor of the British psychologist Richard Gregory who pointed to the practical lessons embodied in artefacts and tools. In an environment full with tools, children will more easily grasp all kinds of technical concepts simply by playing and learning from trial and error. According to Gregory, words are the superior ‘mind tools’, which enable us to transmit more information than any other tool.

A closer look shows that Dennett’s schematic model still reveals a behavioristic bias. Not only does he try to downplay all kinds of innate cognitive mechanisms, he stresses that it “would be a mistake to invoke sentience wherever we find Skinnerian conditioning”. Thus, like most behaviorists, Dennett thinks that ‘negative reinforcement – the ‘punishment’ that “diminishes the likelihood of a repeated performance” is possible without consciousness, without *actual* pain. In fact, he seems to be thinking that real consciousness does only arise at the level of ‘mental tools’, that it should be reserved for humans – and for dogs, of course, especially Dennett’s own dog (or other dogs that obey him).

In fact, what Dennett does is simply building a ‘*scala naturae’* based on a behavioristic model and adding two extra layers, one supplied by the cognitive revolution, and the other supplied by the discovery of culture in animals. Meanwhile, he simply ignores many achievements of ethology, sociobiology and evolutionary psychology. Ethology, for example, has stressed that organisms are packed with innate, specialized learning programs. Animal minds are by no means blank slates, but are more similar to situation-specific, specially designed contact forms or questionnaires, which have to be filled in by experience in a particular environment. Imprinting is the archetype of this kind of knowledge: it works only in the right environment in which a very specific input can be expected. It is a pity that Dennett seems to have missed Lorenz’s ‘*Die Rückseite des Spiegels’* (1973), because he might have added *‘Lorenzian creatures*’ to do more justice to such specialized, custom-made learning mechanisms.

In my opinion, the category of Popperian creatures is also somewhat misleading. Of course, one can compare the internal maps needed by mobile organisms with Popper's hypotheses. But the whole point of making maps is the ability to make choices. Dennett tries to present the emotions as bodily reactions (in the tradition of James-Lange), but there are good reasons to treat them as *cognitive* evaluations that only make sense within the context of the organism's behavioral program. The computer scientist G.E. Pugh wrote an intriguing book about the similarity of the mind with artificial decision-systems in which not only information is processed, but in which decisions are made possible by giving values to all relevant actors, tools, objects and outputs (Pugh, 1978). He shows that the emotions bear striking similarities to such values and interprets the brain as a biological decision-system in which information as such has no meaning whatsoever, apart from the framework of innate evaluations and expectations embodied in the emotions.

It is amazing how much mental and rhetorical gymnastics Dennett needs to bypass the emotions, giving the fact that he admits that even Skinnerian creatures need specialized 'reinforcers' to steer them in the right direction. If only he would admit that pleasure and pain are the most primitive of these reinforcers and that the emotions by which Popperian creatures judge their candidate acts are more than just a 'bodily tribunal' (1996: 119). Apparently, he still has a lot of difficulty in accepting that consciousness may be rooted in the deeper, instinctive parts of the brain and may not be based exclusively on his favorite 'mind tools', but in the first place on the biological necessity to make informed decisions.

The same story goes for the social dimension. For Dennett, friends and lovers are simply conspecifics from which you can learn. As a specialist in the philosophy of mind, he seems to miss large parts of the literature on trust and cooperation. Nowhere in his tower will you find empathy as a product of cooperation or self-consciousness as a means to deal with social control by third parties. In my opinion, Dennett conceives brains as specialized learning devices rather than as the interactive devices from which real organisms are on the outlook for safety, food, friends and partners in order to survive and pass on their genes.

To make his model more realistic, you would at least have to introduce new categories, like Lorenzian creatures, with innate learning mechanisms, and Humphreyan creatures, for socially intelligent strategists (Humphrey, 1976). Popperian creatures could be renamed Pughian creatures. It would make sense to make a difference between traditions originating from elementary social learning (as they exist in different kind of animals) and cumulative, representational culture, which is based on an advanced imagination and an improved learning capacity (as they exist in humans).

All in all, Dennett's fears about 'homunculi' watching 'Cartesian theatres' seem to have inspired a relatively sterile model of the mind. As I have claimed, Dennett is not the only 'naturalist' suffering from dualism-anxiety. Since the logical positivists and since Quine, a lot of Anglo-American philosophers have too easily equated (or rather confused) naturalism with physicalism. This is a pity, since Roy Wood Sellars had already offered a promising alternative with his 'evolutionary naturalism' (1922), which tries to do justice to emergent properties as properties of systems. Such emergent properties are precisely what

we would expect if we realize that natural and sexual selection in the end only test the organism as a *whole* interacting with conspecifics and the environment. Only properties which enhance the overall functionality and fitness of the organism stand a chance of passing the strict jury of natural and sexual selection. On the one hand, they have to be hereditary and replicable; on the other hand, they have to constitute real improvements – adaptations.

But in the eyes of physicalists, emergent properties are difficult to explain. This seems to be the reason that Paul Churchland, at least in some of his works, rejects them, while introducing a new category, 'network properties', at the same time (Churchland, 1989). In the end, Churchland would like to 'reduce' all first-person language to third-person language. One wonders why Churchland has such difficulty in accepting first-person perspectives. Why does he consider scientific objectivity contradictory to subjective experiences? Why would he think that naturalism is incompatible with the limited nature of all knowledge and a realistic perspectivism (see below)? To the extent that Churchland really thinks that first-person perspectives can and should be translated into third-person language, he seems to miss the very essence of consciousness, which by its nature is a private, first-person perspective on the world.

In my opinion, both Dennett and Churchland are physicalists who are largely blind for the possibility that consciousness is a biological adaptation. Consciousness seems to be an emergent phenomenon which by, its very nature, isolates organisms in such a way that they are in a unique position to act according to their own interests. Consciousness enables organisms to view the world from their own 'cockpit' in which they *experience firsthand* both the relevance of different kinds of information and their own behavioral options. Consciousness is, in that respect, more similar to a kind of 'organic dashboard' than to a theater in which the spectator has a purely passive role. To some extent, one can even see the distinction between mind and body as a metaphysical reification and misinterpretation of the first and third person perspectives.

All this would mean that we need a different kind of naturalism than the one proposed by Dennett and Churchland. A naturalism in which we ourselves are part of nature, and in which our first-person and second-person perspectives, introspection and empathy, do matter and, in fact, may contribute to our knowledge about consciousness and other minds. What would such a naturalism look like?

Evolutionary epistemology and Pugh's theory of biological decision systems

At this point, it becomes necessary to introduce evolutionary epistemology. A series of philosophers have reflected on the implications of evolution for epistemology: for example Darwin himself (in his notebooks), Spencer, Nietzsche, and Bergson. Konrad Lorenz' treatise on '*Kant's 'a priori' as a biological phenomenon*' (Lorenz, 1941) is, however, the *locus classicus*. Lorenz claims that what Kant called the 'a priori' – a framework for interpreting sense-data that precedes them – is phylogenetically 'a posteriori' – results from experience during many generations. Thus, the possibility of knowledge has to be understood as a

biological adaptation that enables mobile organisms to orient themselves, to find food and partners, and evade enemies and predators. Knowledge no longer constitutes a mystery, but is an adaptation. If our experience of the world would have been completely fabricated and phantasmagorical, we wouldn't have survived.

This insight enables us to transcend both the classical discussions of (a) idealism versus realism and (b) empiricism versus rationalism. (a) Yes, the world which we experience is a subjective phenomenon – but no, it is not *purely* subjective in that it also reveals some real characteristics of the world. (b) Yes, knowledge starts with sense-data, but at the same time: no, it selects, transforms and interprets these data from the very start. Above all, it enables us to transcend the idea that perspectivism implies relativism, as it is found still in Nietzsche (but see Ortega y Gasset, 1963 [1923]). Knowledge is a biological phenomenon with a number of random characteristics, but it has evolved to enable mobile organisms to orient themselves and to cope with reality. Mobile organisms are using physical properties of the environment to navigate through it. Their nervous systems are trained to do this throughout numerous generations and have evolved causal, stereometric and social expectations to do so efficiently. Hence, Kant's statement that "thoughts without content are empty, sensations without concepts are blind" ("Gedanken ohne Inhalt sind leer, Anschauungen ohne Begriffe, sind blind'', Kant, 1781: 81) should be translated in the language of cognitive ethology: organisms need a framework to interpret their sensations and this framework can be build up quickly each generation as a result of innate biases and specialized innate learning mechanisms.

As far as I know, however, classical evolutionary epistemology did not address the question of consciousness. According to Vollmer (1983 [1975]), evolutionary epistemology is based on a 'hypothetical realism' and justifies the claim that there should at least be a 'partial' correspondence between our subjective experiences and the world - our knowledge of the world has to be 'adequate for survival' ('Überlebungsadequät'). But Vollmer does not address the question *why we experience the world in a subjective fashion* in the first place. The way in which we experience the world is not only subjective as a result of the arbitrariness of the physical characteristics of the world used in it and as a result of our limited capacity to extract knowledge from them. There seems to be a more fundamental reason: we have not evolved, in the first place, to acquire knowledge and an adequate model of our environment, but we have evolved to make adaptive decisions which enable survival and reproduction.

Of course, adaptive decisions have to be well-informed. At the same time, however, it might be suspected that animals have to experience the world in a highly biased way in order to ensure that they make the right decisions with respect to food, danger, predators, potential friends and partners… Some types of foods will taste better than others, some situations will seem more dangerous than others, some conspecifics will be experienced as more attractive than others. It is here that we need concepts that are somewhat broader and more fundamental than Dennett's 'reinforcers'. In my opinion, Dennett only uses this

concept in order to avoid psychological categories (or: in order to suggest that he has discovered a way to talk about subjectivity in an objectivistic language).

It seems to me that the missing link between evolutionary epistemology, psychology and the philosophy of Mind is Pugh's theory about value-driven decision systems (G.E. Pugh, *The biological origin of human values,* 1978). As said, Pugh was a computer engineer working on artificial decision systems. At a given moment he discovered that he needed values to enable his programs to make decisions. Above that, the quality of these decisions could be augmented simply by adding not only more information, but also by assigning more values to relevant situations, objects, persons, etcetera. Then, he noted the correspondence of his artificial decision systems with biological decision systems, which also have to be flexible. Both apparently need values: in biological decision systems these values are ultimately based on "valuative sensations". In the words of Pugh: "The innate built-in values are experienced as good or bad valuative sensations, such as tactile pleasure or pain, comfort or discomfort, joy or sorrow, and good or bad taste. These primary human values include both the 'emotions' and what have been traditionally known as 'biological drives'" (Pugh*,* 1978: 30).

Although he believes to solve the mystery of consciousness and the origin of values this way, Pugh does not speculate about the *origin* of consciousness. He simply analyzes the human mind as a sophisticated decision system loaded with an enormous diversity of qualitative different evaluative sensations. Of course, the individual players in this complex orchestra were not hired in one session. Skinnerian 'reinforcers' may well have been the first players. That does not mean that everything boils down to simple punishment and reward. Pugh stresses that there is a fundamental difference between artificial and biological decision systems: "… In artificial systems, the motivating values are distinguishable from each other only by their sign, magnitude, and timing. In biological systems… the motivating values are typically presented to the conscious mind in qualitatively different forms. For example, thirst is subjectively different from hunger. Pain resulting from a burn on the finger is qualitatively distinguishable from pain resulting from a burn on the elbow. Thus the motivating values are delivered to the conscious mind of specific 'drives' or 'urges' that can be easily distinguished from one another" (Pugh, 1978: 109). According to Pugh, one of the reasons for this is that the different motivating values evolved separately, each coupled to its own specific drive or with the specific type of activity that they have to inspire. In organisms with a strongly developed prefrontal cortex, it is also easier to manage complex decisions when the value consequences of alternatives are easy to distinguish. "The use of distinguishably different values makes it easier to associate specific value components with specific causal factors" (Pugh, 1978: 110).

From the viewpoint of evolutionary psychology it is also easy to see why the motiving values in a biological decision system have to be organized around subjective points of view. In this way, the interests of the individual 'survival machine' are simply woven in their subjective perspectivistic experience. Thus, the way in which we experience the world is a product of selection. Kant's '*Erscheinung',* Schopenhauer's '*Vorstellung*', Heidegger's '*Dasein'* are all products of evolution. The subjective window on the world which is embodied in each

biological decision system is tested by natural and sexual selection each generation. What results is a biological cockpit from which the surrounding world is viewed as a world of opportunities and dangers, in which one's goals have to be achieved and one's interests have to be defended.

Interpreted this way, evolutionary epistemology not only 'justifies' knowledge (as adaptation), but 'criticizes' it as well, albeit in a clearly non-Kantian way. The subjective nature of knowledge does not only result from our limited point of view and from our limited computational power. Foremost, it reflects the fact that knowledge has to send us in particular directions, has to motivate us to behave in certain ways, and has to enable us to make choices between different alternatives. It is to be expected that different organisms, and organisms for example with a different sex or age, will experience the world in different ways, as their experience reflects different options and pitfalls.

This means at the same time, that social knowledge somehow has to deal with these natural differences in interest and perspective. Physicalistic naturalists are often very suspicious about 'folk psychology' – as if it is a completely retarded and prescientific understanding of psychological causality. From the viewpoint of evolutionary epistemology it might be claimed, however, that 'folk psychological' categories may have evolved to refer to the causal factors that really matter. When survival in the real jungle requires adequate physical categories, survival in the *social* jungle could require to some extent adequate psychological categories. Empathy doesn't make much sense when it is based on pure projection. From an evolutionary epistemological point of view it wouldn't be strange to suspect that at least in some mammals, mothers 'understand' their offspring, some dominant individuals 'understand' something about subdominants, males understand something about females, and vice versa. The first person perspective probably not only limits our knowledge about the world, it may also offer unique possibilities to understand other minds.

The evolution of emotions from a more primitive system of punishment and rewards

All in all, consciousness seems to have evolved to motivate us and guide us - even force us - through life. Innate values - in the form of evaluative experiences - are assigned to incoming information in order to enable organisms to weigh their alternatives. What kind of organisms do exhibit such evaluative experiences and in what kind of organisms did consciousness arise?

As we have seen, during large parts of the twentieth century, the very concept of consciousness caused uncomfortable feelings within the scientific community. Many scientists and even philosophers of mind tended to dismiss the concept, often confusing it with dualistic or supernaturalistic interpretations. There is a difference, however, between first and third person *perspectives* and their metaphysical reification. Progress in neurobiology forces us to accept that consciousness and the working brain are two aspects of the same. Like perspectivism and realism, pluralism of experience and metaphysical monism, go hand in hand.

Obvious as this may sound, it seems to have been the relative difficulty of this idea that resulted in the biggest mistake in the history of psychology: the misunderstanding of conditioning. Behaviorists misunderstood operant conditioning, because they thought of it as a kind of purely mechanistic process which can be understood without psychological concepts: the ambivalence of the term 'reinforcement' did it all. Instead, conditioning and 'reinforcement' seem to be based on the most elementary forms of consciousness: pleasure and pain. Pleasure and pain are probably the most elementary and atavistic forms in which psychological feedback is given to desirable or less desirable behavior. Pleasure and pain can only be rewards or punishments when they are experienced consciously.

One way to understand the origin of consciousness is to observe the ongoing relationship between genes and behavior in those groups of organisms in which behavior became gradually more flexible. At some stage during the evolution of mobile organisms, rigid behavioral programs and systems of reflexes became a disadvantage for longer living organisms. Somehow, more autonomy had to be given to the nervous system in order to enable it to adapt better to unpredictable environments. A whole new level of variation and selection was introduced and organisms started to learn via trial and error. But a feedback mechanism was required to enable the nervous system to know the difference between 'right' and 'wrong'. At this point, rewards and punishment evolved as a kind of biological traffic lights. But a reward doesn't work when it isn't felt and the same goes for punishments. Consciousness evolved as the most elementary framework of biological autonomy: on the one hand, genes still determined the general goals of the organism, on the other hand, behavior was no longer directly controlled by genes, but could be adapted to a variety of environments.

If all this is true, there is no need for consciousness to evolve at the moment that behavior is completely automatic. In the terminology of Dennett, 'Darwinian creatures' may be completely unconscious. In principle, all kind of sensors do still not require consciousness. Sensors may activate reflex mechanisms or feed neural networks in which particular input-output relationships are broadly defined. Those input-output relationships can be adjusted to a particular environment by adapting the neural connections via trial and error, apparently without the need for consciousness. Even all kinds of smart devices can 'learn' this way. It is unclear to me to what extent the phenomenon of imprinting can be explained in a similar fashion.

In some organisms the flexibility afforded by this kind of learning must have been insufficient. More autonomy was needed. Organisms had to be allowed to learn by trial and error. A more general feedback system was needed in the form of punishment and rewards. Many thinkers have had the intuition that consciousness somehow is dependent on complex feedback loops. This intuition is compatible with the idea that pleasure and pain are part of a feedback loop in which behavior is either reinforced or not. All higher forms of consciousness may go back to this original system.

Humphrey (*A history of the mind,* 1996) is one of the authors who has reflected on the way in which consciousness could result from a feedback loop. He proposes that

consciousness arises from a feedback loop in which the output fibers are gradually projected back into the brain. As a result, he thinks consciousness is dependent on the sensory cortex. Insects and other invertebrates don't have a sensory cortex, and therefore it is not very likely that they are conscious (Humphrey, 1996: 214). But Humphrey still doesn't ask himself why conscious sensations would be necessary. Consciousness is more than pure sensitivity: it requires at least two types of sensors: first, external sensors that feed information into the neural network, second, internal sensors that establish the meaning of that information for the organism. The second type of sensor could be named evaluative sensors or interest-sensors. These enable the evaluative experiences which according to Pugh constitute the foundation of both consciousness and values.

Ultimately, conscious experiences arise as a result of the necessity to evaluate the alternative moves an organism can make in reaction to its environment. Successful moves have to be approved, less successful moves have to be disapproved. But this approval doesn't come from the outside: organisms have to feel that some moves are simply not in their own interests while others are. Although, we still do not understand how this approval arises, it probably isn't located on the sensory cortex alone, but is somehow linked to pleasure or reward centers in subcortical regions.

Perhaps consciousness is even more primitive than previously suspected. In recently published research it is suggested that even fruit flies show equivalents of fear (Rood, 2015). The behavioral patterns of insects are sometimes complex enough to suspect that they are more than pure automata. Of course, it is much harder to deny the existence of consciousness in vertebrates. In my opinion, it is even possible to make a kind of 'emotion tree' in which we can see how behavior has become more flexible during the evolution of birds and mammals (see appendix). The idea is that with each new level of flexibility, new emotions have to be introduced to enable organisms to deal with their new possibilities.

Thus, my proposal is to link consciousness to the emotions. Consciousness enables organisms to weigh options. The emotions form a system of innate values which force animals to do things or to make particular choices as a result of the way they experience particular situations (e.g. Johnston, 1999). Emotions work because they are felt, because they force the organism to accept its priorities. Of course, many traditional philosophers have proposed a link between consciousness and reason. Even Dennett falls into this tradition, as he links consciousness to 'mind tools', that is: language. In my opinion, language is a very sophisticated system of sharing information and experiences. It can only have evolved at a point in evolution where there was already subjective information to share. Probably many animals are conscious, but only humans have art and poetry to share their experiences at a cognitive level and express and describe their emotions and link them to the details of complex ecological and social situations. Reason is – in the famous words of Hume – "slave of the passions": as an ability to calculate, it has, on its own, no motivating power, but when it is informed by the emotions, it enables us to expand our knowledge of the world and to weight our options in it.

Metaphors for consciousness and neuronal workspace theory

To what kind of model does that lead us? Can we build a model about the way consciousness is related to a plurality of unconscious processes? Of course, an organization chart or flow chart would probably be the best way to model all complex relationships within the brain. Such models are being developed by different neurobiologists, but are still highly speculative, and sometimes not very informative. At the moment, the best a philosopher can probably do is sorting out the right metaphors. Such metaphors are referred to by Dennett as 'intuition pumps' (Dennett, 1984): they enable us to use knowledge from one domain to shed light on another.

As a student, a combination of evolutionary epistemology, sociobiology and Pugh's work on the biological origin of human values inspired me to conceive what I pretentiously called the 'dashboard theory of consciousness' (Slurink, 1998; 2002; 2014). The metaphor of a dashboard seemed to afford an instructive analogy of the relationship between conscious and unconscious. A dashboard provides the driver with a concise overview of the relevant information and of the real options, without offering transparency with respect to the way in which this information is obtained and transferred. Inspired by Pugh and evolutionary epistemology, I stressed that it is not the function of the brain to collect knowledge, but to make the right decisions. It is only logical that the brain would evolve ways to present us information in such a way that it would improve the decision process. Pugh himself also alluded to the adaptive significance of the way in which information is presented:

> "To be able to make decisions, the system must learn to *predict* the value consequences of alternatives. The use of distinguishably different values makes it easier to associate specific value components with specific causal factors. One of the functions of rational thought is to classify the motivating drives or values and associate them with the specific types of activity they are intended to motivate. By delivering each drive to the conscious mind in such a way that it is separately distinguishable, the association problem is greatly simplified "(Pugh, 1978: 110).

Of course, when I started comparing the way in which we experience the world with a dashboard, most people immediately reacted by pointing out that this idea requires a homunculus behind it. I would then react that the dashboard of experience is curled in such a way that it is both dashboard and the observer, subject and the subjective way in which objects are represented. Mobile organisms are comparable with automobiles, but their drivers are locked up in a globular dashboard, a cockpit, and they are both what is inside and the way in which it is perceived. There is no little guy behind the dashboard, because we ourselves are both the dashboards and the little guy behind the wheel. Above that, the whole concept of autonomy is undermined to some extent when we realize that we ourselves are unable to change the goals that evolution has imposed on us via the design of the dashboard.

Years later, in the nineties, I came across Bernard Baars' book "In the theater of consciousness" (1997). The metaphor of a dashboard and that of a theater have a lot in

common: of course, both have the 'problem' of the observer. Apparently, Bernard Baars didn't even bother to defend his metaphor of mind to philosophers like Dennett who rejected the metaphor of a 'Cartesian theater' in his book on consciousness (Dennett, 1992). More important, however, was that Baars came with a somewhat different reason to suspect that consciousness has only limited access to the mind. Baars reasoned that the mind is comprised of a series of specialized information processing units that at some point have to deliver their result to a central information agency. Consciousness is the 'global workspace' in which information from all the senses, from memory and recognition and interpretation centers, comes together and can be subsequently processed.

It is a pity, though, that Baars' model focusses on cognition only. After all, it is behavior that matters above anything else. Baars explains consciousness as a kind of unifying power in which information first converges – 'on stage' in his theater metaphor – and subsequently diverges as it is "widely disseminated to members of the audience" (Baars, 1997: 43). The function of this bottleneck is integration. Consciousness allows us to pick up relevant information from a wide array of specialized circuits that mostly work independently and parallel.

> "If we think of the brain as a distributed system with millions of specialized abilities, the question becomes *how to mobilize all of the specialized unconscious networks in pursuit of survival and reproduction.* This is presumably why the unconscious society of the brain requires a stage, a spotlight, and a director. Consciousness, in this view, serves to disseminate a small amount of information to a vast unconscious audience in the brain. It is the publicity organ in the society of mind." (Baars, 1997: 45).

But why is such integration necessary? In his final chapter on "The functions of consciousness" Baars refers to behavior, again and again, but he does not even mention the role of the emotions. Without the emotions, however, it is hard to understand how the brain can set priorities and make adaptive choices. After all, as Pugh shows, without emotions the brain of a rational being would be forced to think continually about the goals of its behavior and about the survival value of each particular decision. The emotions constitute evolved heuristic values that enable us to make the choices that matter at each particular stage in our lives. For that reason, I find the metaphor of a dashboard somewhat more informative than the one of a theater, although admittedly both are inadequate and lead to questions about the observer and the person behind the wheel respectively.

The last decade, the theory of Baars has become increasingly popular, and for good reasons. It predicted that consciousness correlates with a large scale integration of activity in diverse regions of the brain, and that is exactly what was demonstrated with the use of fMRI-scanners. The resulting new model was named 'neuronal workspace theory' by the French neuropsychologist Dehaene. Dehaene describes a series of 'signatures of consciousness' with which one can recognize consciousness, even in coma patients. In the future these signatures could even help us to decide which animals are conscious and which ones are not (Dehaene, 2014).

The recent progress in the study of consciousness was not only a result of new technology, like fMRI-scanning. It was also the result of the realization, by Baars and others, that exactly situations in which visual stimuli do *not* penetrate into consciousness allow experimental psychologists to study the differences with situations in which they do. For example, in binocular rivalry, only one of the two images presented to the eyes, can be consciously seen at the same time. Images which are shown too short, do not become conscious, but do nevertheless have an influence on later observations. In this way, it becomes possible to study the differences between observations that do become conscious and those which do not.

fMRI and older techniques (EEG, MEG) are subsequently used to get a picture of what happens in the brain at the moment that something does become conscious. Consciousness displays itself in a typical sequence in which first different brain regions 'ignite' and then a more global 'avalanche' of information processing starts, finally resulting in a 'brain web' of synchronized neuronal oscillations. In the case of vision, this brain web is characterized by bidirectional causal relations from the visual cortex to the frontal lobe, and vice versa, together with a typical pattern of brain waves, both in the lower and higher frequencies.

All this suggests a third metaphor for consciousness: the boardroom. If we compare the multitude of specialized centers in the brain with regular employees in a factory, consciousness is a team meeting of managers in which decisions are made. Most of the time, the workers in different parts of the factory just do their job, but sometimes their tasks have to be coordinated, and the managers are called together in the boardroom. If the decisions made in the boardroom turn out wrong, tensions rise, and new strategies and working plans have to be thought out… Even the directors in the board room don't have the final word: to some extent, they are just puppets controlled by shareholders… (Which brings us back to genes).

Perhaps we can conclude that metaphors can be instructive, but are always inadequate... All three metaphors mentioned suggest that consciousness has a function. Somehow behavior has to be controlled and coordinated in order to adjust it to circumstances. When reflexes and rigid behavioral programs no longer suffice, flexibility is required. Global workspace theory and neuronal workspace theory are plausible models about the way in which coordination is achieved in a brain full of specialized circuits. Although, I personally believe, they still leave out the decisive factor: the emotions.

Affective neuroscience, flexibility and the function of consciousness

Konrad Lorenz once observed that "there are very simple neural processes which are associated with intense experiences and very complex ones, analogue to rational operations, which in spite of that, are beyond experience and even beyond self-observation"(Lorenz, 1973). Although unclear about the nature and location of consciousness, Pugh situates the value system of biological decision systems in the midbrain, especially in the hypothalamus.

> "Evidently, the forebrain houses a 'rational' decision system; the midbrain houses a sensory processor and special-purpose computer that drives the value system; and the hindbrain houses an output processor which converts general decision into specific muscle commands. If this perspective is correct, it suggests that the rudimentary value-decision system is a surprisingly old evolutionary invention." (Pugh, 1978: 134)

Somewhat later, he argues that "the extreme complexity and refinement of the innate human value system may reflect the operation of …the… large frontal lobes" (Pugh, 1978: 135). He reasons that during human evolution the midbrain proved too small and that the frontal lobes and sensory cortex took over some of its tasks.

Although modern research has shown that consciousness is by no means exclusively linked to the cortex, few researchers have looked for its origin in deeper parts of the brain. Often consciousness is still largely linked with rationality and neural activity across the cortex, as Dehaene does to some extent. Modern neuroimaging techniques are based on the assumption that the magnetic changes of oxygenated blood reflect all brain activities, but whether all changes in the midbrain are mapped this way remains unclear. Even the much celebrated but sometimes obscure Damasio (1999) is not very helpful in creating a truly interdisciplinary model of the brain: his work is mainly based on human brains and he hardly refers to animal behavior. For this reason, the new work of Panksepp & Biven on "The archeology of mind" (2012) comes as a revelation. For the first time, the evolutionary anatomy of seven emotional subsystems becomes clear and it is shown that they largely depend on subcortical networks, for example those surrounding the 'periaqueductal gray'.

Panksepp and Biven prove their point by pointing out that human babies, "who are born basically without cerebral hemispheres, can grow up to be affectively vibrant children if they are raised in nurturing and social engaging environments" (Panksepp & Biven, 14). Laboratory animals without cortex are even more emotional than normal animals, suggesting that the cortex rather suppresses and corrects the basic emotions than exclusively housing them. The largest part of their book describes the seven emotional subsystems which all are based in the midbrain and do their work both in animals and humans, often showing surprising continuities and analogies in pathological behavior. They show that human neuroticism and psychopathology can sometimes only be understood within the larger framework of the atavistic emotional systems that we share with other animals, which Panksepp & Biven call SEEKING, RAGE, FEAR, LUST, CARE PANIC/GRIEF, PLAY.

Of course, Panksepp's work is largely based on rats. Despite the fact that he discovered that rats 'laugh' when they are tickled, rats probably do not exhibit the full spectrum of social emotions which can be observed in primates. The emotional systems that enable cooperation and culture have probably somehow evolved out of the more atavistic emotions, described by Panksepp: for example, moralistic aggression or indignation probably evolved from RAGE, trust on which all cooperation is based may have evolved from CARE, and so forth. In essence, evolution does seldomly create things *ex nihilo,* as God once was supposed to do, and the same probably goes for emotions. At the same time, it is clear that the emotional systems, which, according to Panksepp, are shared by all mammals, didn't

stop evolving in intelligent groups like cetaceans, elephants, apes, and hominids. In some groups of social animals, for example, there are links between group size and the size of particular parts of the brain (Dunbar, 1993; 2014).

For this reason, it should be possible to draw a phylogenetic tree of emotions, which corresponds to the behavioral repertoires of the animals exhibiting them (see appendix). If I am right in following the intuitions of Lorenz, Pugh and Panksepp and the emotions 'drive' flexible behavior via consciousness, this tree reflects the way in which the requirement of flexible behavior drove psychological complexity in vertebrates. Of course, this tree of emotions should not be understood as a hero story, or a story of progress: it is the story of a specialization, a specialization that in one specialized group, the hominids, led to our unique intellectual and moral capacities (Suddendorf, 2013). It is important to realize that the way in which we express our emotions is phylogenetically contingent. Paul Ekman's (2003) classification of the six basic emotions in humans, based on the universal expression of these emotion in humans, have to be seen as one of the many branches of the tree. Of course, any attempt at drawing that tree is, at this moment, a speculative exercise. At the same time, the need for a phylogeny of emotions arises naturally from the realization that even apparently uniquely human properties like morality and cumulative culture have their roots in much older behavioral and cognitive drives.

Conclusion

Consciousness is an evolved property which transforms information into a uniquely perspectivistic, first person experience of the world. In this way, it does not only represent the outside world, but also the unique interests of the individual. The idea that it is simply an epiphenomenon is therefore unlikely: instead, it seems to enable flexible decisions in which a plurality of factors is weighed, as Pugh proposed with his theory of the value driven decision systems. That doesn't mean that all decision processes are conscious: instead, consciousness seems to enable flexible decisions on the basis of a weighing within a *limited overview* of the available information. This can be expressed by a 'dashboard' model of consciousness, which bears similarities to the 'workspace model' of Bernard Baars. The emotions constitute the 'value system' which makes decisions possible: as Lorenz and Pugh suspected and Panksepp proved, they have their basis in older parts of the midbrain, although the elaborate value system underlying the unique cognitive and cooperative abilities of humans also require parts of the forebrain and of the cortex. For this reason, it is possible to draw a speculative 'tree of emotions' which shows their phylogenetic history in some groups of vertebrates: this tree, if designed properly, would illustrate that the emotions underlying cooperation and even morality, are based on an evolutionary older system which is much more widespread in the animal kingdom. This system may have evolved from a more primitive punishment and reward system coupled to conditioning. If this is true, the ability to 'shape' behavior through 'conditioning' and 'reinforcement' - as described by the behaviorists - would in fact be an indication of consciousness – *contrary* to the interpretation of most behaviorists.

Acknowledgements. The author thanks Andreas Blank, Brenda de Groot, Edwin van Leeuwen and Bram Slurink for their helpful comments.

Appendix: a speculative tree of emotions

Context	Emotion		Animals
1. Body	Pain ←→ pleasure	Atavistic affects	Fish, Amphibians, Reptiles?
2. Safety → 5	FEAR , *FEAR* ←→ wellbeing		
2. Food	Hunger ←→ satisfaction *DISGUST* ←→ appetite		
4. Reproduction → 6	LUST, desire		
5. Exploration, Learning	SEEKING, curiosity, excitement ←→ boredom ➔ *SURPRISE*	Elementary emotions	Birds? Mammals
6. Breeding care & Attachment → 8, 11	CARE, Love & concern ←→ PANIC, loss		
7. Safety in numbers , social learning → 9	PLAY, sociability ←→ loneliness		
8. Competition for Partners → 10	Jealousy←→ infatuation ➔ love ←→ hatred		
9. Competition for dominance → 10, 11, 12	RAGE, *ANGER*, pride, self-confidence, arrogance ←→ submissiveness ➔ Admiration		
10. Good & bad luck, winning, losing	*HAPPINESS ←→ SADNESS*		
11. Cooperation → 12	Trust ←→ distrust, reproach, indignation ←→ guilt, duty, gratitude	Social emotions	Long living intelligent mammals
12. Sympathy + time travel	Empathy, compassion, malicious pleasure, sadism ➔ shame, regret, remorse, guilt		Apes, hominins, Man

A philosopher's speculation about the evolution of emotions, based on behavioral observation and informed guessing. An attempt is made to include both Panksepp's seven emotional systems (Seeking, Lust, Care, Rage, Fear, Panic, Play), based on rats, and Ekman's basic emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise), largely based on humans. Arrows in the left column (→) refer to hypothetic phylogenetic relationships ('evolves into' or 'enables the evolution of'). Double-headed arrows (←→) in the second column refer to bipolar affects or emotions.

References

- Baars, B.J. 1997. *In the Theater of Consciousness; The Workspace of the Mind.* Oxford: Oxford UP.
- Bergson, H. 1908. L'évolution créatrice. Paris: F. Alcan.
- Bowlby, J. 1980. Attachment and loss. Vol. 3. New York: Basic books.
- Chadwick, H. 1980. Origen: Contra Celsum. Cambridge University Press, 1980.
- Churchland, P. 1989. *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: The MIT Press.
- Damasio, A. 1994. *Descartes' Error — Emotion, Reason and the Human Brain*. New York Putnam Publishing.
- Darwin, F. 1892 [1958] Charles Darwin, Autobiography. Mineola: Dover publications.
- Dehaene, S. 2014. *Consciousness and the brain. Deciphering how the brain codes our thoughts.* New York: Viking.
- Dennett, D.C. 1984. *Elbow Room: the varieties of free will worth wanting.* Oxford: Oxford UP.
- Dennett. D.C. [1996] 2001. *Kinds of minds. The origins of consciousness.* London: Phoenix.
- Dunbar, R.I.M. 1993. The co-evolution of neocortical size, group size and language in humans. *Behavioural and Brain Sciences*, 16, 681-735.
- Dunbar, R. 2014. Human evolution: A Pelican introduction. Harmondsworth: Penguin.
- Ekman, P. 2003. Emotions revealed. New York: Henry Holt.
- Humphrey, N. K. (1976). The social function of the intellect. In P. P. G. Bateson & R. A. Hinde (eds.). Growing points in ethology. Cambridge: Cambridge University Press
- Humphrey, N. 1992. *A History of the Mind. Evolution and the birth of consciousness.*New York: Copernicus/Springer.
- Johnston, V.S. 1999. *Why we feel. The science of human emotions.* Reading, MA: Perseus Books.
- Kant, I. 1981. Critik der reinen Vernunft, Riga: Hartknoch, 1781 (http://www.deutschestextarchiv.de/book/show/kant_rvernunft_1781)
- Lorenz, K. 1941. Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie. In: K. Lorenz & F. Wuketits (Hrsg.), 1983. *Die Evolution des Denkens,* Zürich/München: Piper & Co.
- Lorenz, K. 1973. *Die Rückseite des Spiegels*, München: Piper.
- Ortega y Gasset, J. 1963 [1923]. *De taak van onze tijd (The task of our time)*, Den Haag (*El tema de nuestro tiempo*, Madrid, 1923.)
- Panksepp, J. & L. Biven. 2012. The Archeology of Mind. Neuroevolutionary Origins of Human Emotions. New York: Norton.
- Pugh, G.E. 1978. *The biological origin of human values.* London & Henley: Routledge & Kegan Paul.
- Rood, J. 2015. Fruit fly fear. *The scientist,* May 15.
- Sellars, R.W. 1922. *Evolutionary Naturalism,* Chicago & London: Open Court Publishing Company.
- Sellars, R. W. 1973. A Naturalistic Theory of Value and Valuation & Agential Causality and Free Will. *Neglected Alternatives; Essays by Roy Wood Sellars.* W.P. Warren, ed. Lewisburg: Bucknell UP.
- Sieben, H. 1977. ORIGENES, Vier Bücher von den Prinzipien, hrsg. v. Herwig Görgemanns u. Heinrich Karpp. Darmstadt: Wissenschaftliche Buchgesellshaft.
- Slurink, P. 1998. Why consciousness evolved. In: Ton Derksen (ed.), *The Promise of Evolutionary Epistemology*. Tilburg: Tilburg University Press.
- Suddendorf, T. 2013. *The Gap. The science of what separates us from other animals.* New York: Basic Books.
- Slurink, P. 2002. *Why some apes became humans.* Nijmegen: Dissertation. (Free pdf at http://hdl.handle.net/2066/32105).
- Slurink, P. 2014. *Aap zoekt zin; waarom wij bewustzijn, vrije wil, cultuur en religie hebben,* Leusden: ISVW Publishers.
- Vollmer G. [1975] 1983[3]. *Evolutionäre Erkenntnistheorie,* Stuttgart: Heinzel Verlag.
- Vollmer, G. 1988. *Was können wir wissen? Band 1: Die Natur der Erkenntnis; Band 2: Die Erkenntnis der Natur.* Stuttgart: Hirzel Verlag.

References

Brugger, E. C. (2013) "D. Alan Shewmon and the PCBE's White Paper on Brain Death", *Journal of Medicine and Philosophy* 38, pp. 205–218.

Halevy, A. (2001) "Beyond Brain Death?" *Journal of Medicine and Philosophy* 26, pp. 493–501.

Noonan, H. (1989) *Personal Identity*, London: Routledge.

Olson, E. T. (1997) *The Human Animal: Personal Identity Without Psychology*, New York: Oxford University Press.

— (2003) "An Argument for Animalism", in *Personal Identity*, edited by R. Martin & J. Barresi, Oxford: Blackwell, pp. 318–334.

— (2013) "The Person and the Corpse", in *The Oxford Handbook of Philosophy of Death*, edited by B. Bradley et al., Oxford: Oxford University Press, pp. 80–96.

Shewmon, D. A. (2001) "The Brain and Somatic Integration", *Journal of Medicine and Philosophy* 26, pp. 457–478.

Tzinman, R. (2016) "Against the Brainstem View of the Persistence of Human Animals", this volume.

van Inwagen, P. (1990) *Material Beings*, Ithaca: Cornell University Press.

13
Consciousness as an Adaptation— What Animals Feel & Why

Pouwel Slurink

> While the behaviorists recognized that aversive UCS (Unconditioned Stimulus), such as predator odor or electric shock, can serve as 'punishments' in many learning tasks, they could overlook as irrelevant that fact that UCRs, such as fearfulness, also have an internal feel to them. Other UCS could serve as 'rewards' that would promote the learning of approach behaviors rather than avoidance behaviors. There has traditionally been little discussion, however, on any corresponding feelings underlying the logic of behavioral learning in animals. Of course, it is likely that rewards and punishments only work so well to control learning because they generate affective feelings in the brain. The spooky process of *reinforcement* may reflect the way feelings work in the brain.[1]

Introduction

Generally speaking, most western philosophers have had considerable difficulty in conceptualizing different 'kinds of minds'[2]. The main cause seems to be a preoccupation with our own special position and responsibility and the dominance of a dualistic, religious Platonic-Christian interpretation thereof. In Platonism and Hinduism a dualistic belief in reincarnation does not exclude a gradualism with respect to the relation between man and the other animals. In early Christianity, however, the uniqueness of each individual human soul is stressed at the expense of the experience of continuity and connectedness with other living beings. The notions of an exclusive relationship between man and (a transcendent) God, and of a largely divine but in part human mission to change the world,

[1] Panksepp & Biven (2012), 12.
[2] Dennett (1996).

tipped the balance strongly in favor of anthropocentrism. Of course, one of the advantages of this position was that it enabled their proponents to defend (their) morality as a God-given, absolute phenomenon.

The transition from ancient to Christian thinking about animals is beautifully documented in Origen's *Contra Celsum*[3]. As a typical ancient thinker, Celsus attributes an enormous memory to elephants, linguistic talents to ants, and foresight to birds. In contrast, Origen makes an absolute distinction between animals and 'rational souls'. Only 'rational souls' are sufficiently disembodied to be able to move from body to body. In his *Peri Archon*[4] he presents the material world as a kind of dungeon in which fallen angels and other rational souls can find their way back to righteousness. That implies, at the same time, that the world was created exclusively *for* those rational souls and their bliss.

Of course, Origen was considered a heretic and not all western philosophers were as extreme in their opinions on the differences between animals and man as he was. Yet, the identification of mind-body dualism with a dualism between man and other animals is repeated time and again in Western philosophy. The same goes for the notion that man is a unique being with the unique ability to transcend his own impulses and even all natural causality—think especially of Descartes and Kant. Even the 'evolutionary' philosopher Bergson, with his *'Évolution créatrice'*[5], still tries to find holes in the chains of necessity and to unite evolution, spirituality, human uniqueness, free will, and moral progress ('Dieu se fait').

Darwin put us back with two (not four) feet on the ground. His 'theory of descent with modification through natural selection' enables us to find a more balanced view of the relationship between man and other species. In contrast to Lamarck's theory of evolution, Darwin's theory does not imply mysterious 'nerve energies' and a disguised notion of progress. After his trip around the world on the Beagle, the young Darwin became convinced of the 'transmutation of species' and decided to search for a completely

[3] Circa 248 a.d.; See Chadwick (1980).
[4] Circa 212–215 a.d.; See Origenes (1985).
[5] Bergson (1908).

naturalistic—almost geological—mechanism behind it. After several years, he "happened to read for amusement" Malthus' treatise on population[6]. He immediately realized that the notion of a 'struggle for life' resulting from overpopulation was exactly what he had been looking for all the time. He realized that, when this struggle goes on in all life and is combined with tiny hereditary differences among individuals, this would result in the survival of those varieties which are slightly better adapted to the current circumstances than others. This simple combination of reproduction, variation and selection would then be enough to explain 'the transmutation of species'.

Why is this insight so important? Firstly, it gives us a concise explanation of evolution that links biology, and even psychology, to the rest of the natural sciences. Secondly, it indicates that many properties of both animals and humans are either adaptations or result from them. New features, including those that constitute human uniqueness, do not come out of the blue, but arise from necessity: mutants with them somehow fare better than individuals without them. Thus, if man is unique—as is each species—his unique properties must have originally constituted an advantage above other properties in our ancestral environments[7].

Thus, instead of simply stating that man is 'just another ape', Darwin's theory implies that we can rephrase discussions about human uniqueness and about the differences between man and other animals in terms of adaptations that are either shared—as a result of common evolutionary pathways—or are uniquely human—as a result of a uniquely human evolutionary trajectory. This way, a much more nuanced view about the talents of different kinds of animals is possible as well as a deeper understanding of human uniqueness. Of course, this kind of understanding requires an enormous knowledge about the conditions in which apes, hominids and humans evolved and to which they are adapted to some extent. Although much progress has been made in this area, a series of mysteries is likely to remain here as well.

[6] In July, 1838; See Darwin (1958).
[7] 'The environment of evolutionary adaptedness'; See Bowlby (1980).

Reductionist naturalism: from dualism-anxiety to naive objectivism

Of course, the Darwinian theory of evolution has been around for a while and there have been many attempts at deriving models of animal and human cognition from it. During the twentieth century, however, both science and philosophy were still often too much dominated by opposing metaphysical views to be able to appreciate or even understand the advantages of a Darwinian approach. On the one side, positivists and physicalistic naturalists tended to view the very concept of 'consciousness' as a remnant of dualistic metaphysics—hence the popularity of behaviorism with its dogmatic rejection of both empathy and introspection as sources of knowledge. On the other side, phenomenological and hermeneutical philosophers tended to dismiss naturalism altogether in the name of pure 'experience' or, at a later stage, some version of sophisticated relativism. Only a small collection of independent thinkers, like Roy Wood Sellars and Konrad Lorenz, managed to keep a cool head during these heated debates and managed to transcend the presuppositions leading to its false oppositions. During the computer age, behaviorism was finally succeeded by cognitivism, but this again led to a one-sided model of the mind as a purely passive 'information processing' organ.

As a result, even philosophers that think of themselves as naturalists and Darwinists are sometimes still producing models of cognitive evolution that do not address the question of functionality adequately. Even 'naturalists' like Dennett and Churchland suffer from a dualism-anxiety which evokes memories of the logical positivists. Let me illustrate this with Dennett's model of the 'various design options for brains', which is in fact a kind of rational reconstruction of cognitive evolution. Somewhat pretentious and bombastic Dennett speaks of 'the Tower of Generate-and-Test'[8].

According to Dennett, there are about four floors in this cognitive tower and each new floor reflects a new level of behavioral sophistication. "As each new floor of the Tower gets constructed, it empowers the organisms at that level to find better and better

[8] Dennett (1996).

moves, and find them more efficiently". At the ground floor we find *Darwinian creatures,* which are completely 'field-tested' and unable to correct their own behaviour during their short lives. Any small error in their innate behavioral programs may lead to their premature death.

The next floor is inhabited by creatures with built-in 'reinforcers' that enable them to learn 'smart moves' by a "long, steady process of training and shaping by the environment''. "These individuals ... confronted... the environment by generating a variety of actions, which they tried out, one by one, until they found one that worked". Dennett calls them *Skinnerian creatures* and claims that Hume's associationism was a predecessor of both behaviorism and connectionism.

The next floor is reserved for what Dennett calls *Popperian creatures,* which have some kind of internal maps or representation of the world in which they can try out their behavioral experiments virtually. Strikingly, but perhaps characteristically, Dennett evades notions of 'innate ideas' in this context, or of the 'a priori'. The only thing he deems necessary are "perceptual mechanisms designed to ignore most of the flux of stimuli", which "concentrate on the most useful, most reliable information".

Dennett's last floor is reserved for organisms that learn from one another. Dennett calls them *Gregorian creatures,* in honor of the British psychologist Richard Gregory who pointed to the practical lessons embodied in artefacts and tools. In an environment full with tools, children will more easily grasp all kinds of technical concepts simply by playing and learning from trial and error. According to Gregory, words are the superior 'mind tools', which enable us to transmit more information than any other tool.

A closer look shows that Dennett's schematic model still reveals a behavioristic bias. Not only does he try to downplay all kinds of innate cognitive mechanisms, he stresses that it "would be a mistake to invoke sentience wherever we find Skinnerian conditioning". Thus, like most behaviorists, Dennett thinks that 'negative reinforcement—the 'punishment' that "diminishes the likelihood of a repeated performance" is possible without consciousness, without *actual* pain. In fact, he seems to be thinking that real consciousness does only arise at the level of 'mental tools', that it

should be reserved for humans—and for dogs, of course, especially Dennett's own dog (or other dogs that obey him).

In fact, what Dennett does is simply building a '*scala naturae*' based on a behavioristic model and adding two extra layers, one supplied by the cognitive revolution, and the other supplied by the discovery of culture in animals. Meanwhile, he simply ignores many achievements of ethology, sociobiology and evolutionary psychology. Ethology, for example, has stressed that organisms are packed with innate, specialized learning programs. Animal minds are by no means blank slates, but are more similar to situation-specific, specially designed contact forms or questionnaires, which have to be filled in by experience in a particular environment. Imprinting is the archetype of this kind of knowledge: it works only in the right environment in which a very specific input can be expected. It is a pity that Dennett seems to have missed Lorenz's '*Die Rückseite des Spiegels*' (1973), because he might have added '*Lorenzian creatures*' to do more justice to such specialized, custom-made learning mechanisms.

In my opinion, the category of Popperian creatures is also somewhat misleading. Of course, one can compare the internal maps needed by mobile organisms with Popper's hypotheses. But the whole point of making maps is the ability to make choices. Dennett tries to present the emotions as bodily reactions (in the tradition of James-Lange), but there are good reasons to treat them as *cognitive* evaluations that only make sense within the context of the organism's behavioral program. The computer scientist G.E. Pugh wrote an intriguing book about the similarity of the mind with artificial decision-systems in which not only information is processed, but in which decisions are made possible by giving values to all relevant actors, tools, objects and outputs[9]. He shows that the emotions bear striking similarities to such values and interprets the brain as a biological decision-system in which information as such has no meaning whatsoever, apart from the framework of innate evaluations and expectations embodied in the emotions.

[9] Pugh (1978).

It is amazing how much mental and rhetorical gymnastics Dennett needs to bypass the emotions, giving the fact that he admits that even Skinnerian creatures need specialized 'reinforcers' to steer them in the right direction. If only he would admit that pleasure and pain are the most primitive of these reinforcers and that the emotions by which Popperian creatures judge their candidate acts are more than just a 'bodily tribunal'[10]. Apparently, he still has a lot of difficulty in accepting that consciousness may be rooted in the deeper, instinctive parts of the brain and may not be based exclusively on his favorite 'mind tools', but in the first place on the biological necessity to make informed decisions.

The same story goes for the social dimension. For Dennett, friends and lovers are simply conspecifics from which you can learn. As a specialist in the philosophy of mind, he seems to miss large parts of the literature on trust and cooperation. Nowhere in his tower will you find empathy as a product of cooperation or self-consciousness as a means to deal with social control by third parties. In my opinion, Dennett conceives brains as specialized learning devices rather than as the interactive devices from which real organisms are on the outlook for safety, food, friends and partners in order to survive and pass on their genes.

To make his model more realistic, you would at least have to introduce new categories, like Lorenzian creatures, with innate learning mechanisms, and Humphreyan creatures, for socially intelligent strategists[11]. Popperian creatures could be renamed Pughian creatures. It would make sense to make a difference between traditions originating from elementary social learning (as they exist in different kind of animals) and cumulative, representational culture, which is based on an advanced imagination and an improved learning capacity (as they exist in humans).

All in all, Dennett's fears about 'homunculi' watching 'Cartesian theatres' seem to have inspired a relatively sterile model of the mind. As I have claimed, Dennett is not the only 'naturalist' suffering from dualism-anxiety. Since the logical positivists and since Quine, a lot of Anglo-American philosophers have too easily

[10] Dennet (1996), 119.
[11] Humphrey (1976).

equated (or rather confused) naturalism with physicalism. This is a pity, since Roy Wood Sellars had already offered a promising alternative with his 'evolutionary naturalism'[12], which tries to do justice to emergent properties as properties of systems. Such emergent properties are precisely what we would expect if we realize that natural and sexual selection in the end only test the organism as a *whole* interacting with conspecifics and the environment. Only properties which enhance the overall functionality and fitness of the organism stand a chance of passing the strict jury of natural and sexual selection. On the one hand, they have to be hereditary and replicable; on the other hand, they have to constitute real improvements—adaptations.

But in the eyes of physicalists, emergent properties are difficult to explain. This seems to be the reason that Paul Churchland, at least in some of his works, rejects them, while introducing a new category, 'network properties', at the same time[13]. In the end, Churchland would like to 'reduce' all first-person language to third-person language. One wonders why Churchland has such difficulty in accepting first-person perspectives. Why does he consider scientific objectivity contradictory to subjective experiences? Why would he think that naturalism is incompatible with the limited nature of all knowledge and a realistic perspectivism (see below)? To the extent that Churchland really thinks that first-person perspectives can and should be translated into third-person language, he seems to miss the very essence of consciousness, which by its nature is a private, first-person perspective on the world.

In my opinion, both Dennett and Churchland are physicalists who are largely blind for the possibility that consciousness is a biological adaptation. Consciousness seems to be an emergent phenomenon which by, its very nature, isolates organisms in such a way that they are in a unique position to act according to their own interests. Consciousness enables organisms to view the world from their own 'cockpit' in which they *experience firsthand* both the relevance of different kinds of information and their own behavioral options. Consciousness is, in that respect, more similar to a kind

[12] Sellars (1922).
[13] Churchland (1989).

of 'organic dashboard' than to a theater in which the spectator has a purely passive role. To some extent, one can even see the distinction between mind and body as a metaphysical reification and misinterpretation of the first and third person perspectives.

All this would mean that we need a different kind of naturalism than the one proposed by Dennett and Churchland. A naturalism in which we ourselves are part of nature, and in which our first-person and second-person perspectives, introspection and empathy, do matter and, in fact, may contribute to our knowledge about consciousness and other minds. What would such a naturalism look like?

Evolutionary epistemology and Pugh's theory of biological decision systems

At this point, it becomes necessary to introduce evolutionary epistemology. A series of philosophers have reflected on the implications of evolution for epistemology: for example Darwin himself (in his notebooks), Spencer, Nietzsche, and Bergson. Konrad Lorenz' treatise on '*Kant's 'a priori' as a biological phenomenon*'[14] is, however, the *locus classicus*. Lorenz claims that what Kant called the 'a priori'—a framework for interpreting sense-data that precedes them—is phylogenetically 'a posteriori'—results from experience during many generations. Thus, the possibility of knowledge has to be understood as a biological adaptation that enables mobile organisms to orient themselves, to find food and partners, and evade enemies and predators. Knowledge no longer constitutes a mystery, but is an adaptation. If our experience of the world would have been completely fabricated and phantasmagorical, we wouldn't have survived.

This insight enables us to transcend both the classical discussions of (a) idealism versus realism and (b) empiricism versus rationalism. (a) Yes, the world which we experience is a subjective phenomenon—but no, it is not *purely* subjective in that it also reveals some real characteristics of the world. (b) Yes, knowledge starts with sense-data, but at the same time: no, it selects, trans-

[14] Lorenz (1941).

forms and interprets these data from the very start. Above all, it enables us to transcend the idea that perspectivism implies relativism, as it is found still in Nietzsche[15]. Knowledge is a biological phenomenon with a number of random characteristics, but it has evolved to enable mobile organisms to orient themselves and to cope with reality. Mobile organisms are using physical properties of the environment to navigate through it. Their nervous systems are trained to do this throughout numerous generations and have evolved causal, stereometric and social expectations to do so efficiently. Hence, Kant's statement that "thoughts without content are empty, sensations without concepts are blind"[16] should be translated in the language of cognitive ethology: organisms need a framework to interpret their sensations and this framework can be build up quickly each generation as a result of innate biases and specialized innate learning mechanisms.

As far as I know, however, classical evolutionary epistemology did not address the question of consciousness. According to Vollmer[17], evolutionary epistemology is based on a 'hypothetical realism' and justifies the claim that there should at least be a 'partial' correspondence between our subjective experiences and the world—our knowledge of the world has to be 'adequate for survival' ('Überlebensadäquat'). But Vollmer does not address the question *why we experience the world in a subjective fashion* in the first place. The way in which we experience the world is not only subjective as a result of the arbitrariness of the physical characteristics of the world used in it and as a result of our limited capacity to extract knowledge from them. There seems to be a more fundamental reason: we have not evolved, in the first place, to acquire knowledge and an adequate model of our environment, but we have evolved to make adaptive decisions which enable survival and reproduction.

Of course, adaptive decisions have to be well-informed. At the same time, however, it might be suspected that animals have to

[15] See Ortega y Gasset ([1923] 1963).
[16] "Gedanken ohne Inhalt sind leer, Anschauungen ohne Begriffe, sind blind''; See Kant (1781), 81.
[17] Vollmer ([1975] 1983).

experience the world in a highly biased way in order to ensure that they make the right decisions with respect to food, danger, predators, potential friends and partners... Some types of foods will taste better than others, some situations will seem more dangerous than others, some conspecifics will be experienced as more attractive than others. It is here that we need concepts that are somewhat broader and more fundamental than Dennett's 'reinforcers'. In my opinion, Dennett only uses this concept in order to avoid psychological categories (or: in order to suggest that he has discovered a way to talk about subjectivity in an objectivistic language).

It seems to me that the missing link between evolutionary epistemology, psychology and the philosophy of Mind is Pugh's theory about value-driven decision systems[18]. As said, Pugh was a computer engineer working on artificial decision systems. At a given moment he discovered that he needed values to enable his programs to make decisions. Above that, the quality of these decisions could be augmented simply by adding not only more information, but also by assigning more values to relevant situations, objects, persons, etcetera. Then, he noted the correspondence of his artificial decision systems with biological decision systems, which also have to be flexible. Both apparently need values: in biological decision systems these values are ultimately based on "valuative sensations". In the words of Pugh: "The innate built-in values are experienced as good or bad valuative sensations, such as tactile pleasure or pain, comfort or discomfort, joy or sorrow, and good or bad taste. These primary human values include both the 'emotions' and what have been traditionally known as 'biological drives'"[19].

Although he believes to solve the mystery of consciousness and the origin of values this way, Pugh does not speculate about the *origin* of consciousness. He simply analyzes the human mind as a sophisticated decision system loaded with an enormous diversity of qualitatively different evaluative sensations. Of course, the individual players in this complex orchestra were not hired in one session. Skinnerian 'reinforcers' may well have been the first players. That does not mean that everything boils down to simple pun-

[18] Pugh (1978).
[19] Pugh (1978), 30.

ishment and reward. Pugh stresses that there is a fundamental difference between artificial and biological decision systems: "... In artificial systems, the motivating values are distinguishable from each other only by their sign, magnitude, and timing. In biological systems... the motivating values are typically presented to the conscious mind in qualitatively different forms. For example, thirst is subjectively different from hunger. Pain resulting from a burn on the finger is qualitatively distinguishable from pain resulting from a burn on the elbow. Thus the motivating values are delivered to the conscious mind of specific 'drives' or 'urges' that can be easily distinguished from one another"[20]. According to Pugh, one of the reasons for this is that the different motivating values evolved separately, each coupled to its own specific drive or with the specific type of activity that they have to inspire. In organisms with a strongly developed prefrontal cortex, it is also easier to manage complex decisions when the value consequences of alternatives are easy to distinguish. "The use of distinguishably different values makes it easier to associate specific value components with specific causal factors"[21].

From the viewpoint of evolutionary psychology it is also easy to see why the motiving values in a biological decision system have to be organized around subjective points of view. In this way, the interests of the individual 'survival machine' are simply woven in their subjective perspectivistic experience. Thus, the way in which we experience the world is a product of selection. Kant's '*Erscheinung*', Schopenhauer's '*Vorstellung*', Heidegger's '*Dasein*' are all products of evolution. The subjective window on the world which is embodied in each biological decision system is tested by natural and sexual selection each generation. What results is a biological cockpit from which the surrounding world is viewed as a world of opportunities and dangers, in which one's goals have to be achieved and one's interests have to be defended.

Interpreted this way, evolutionary epistemology not only 'justifies' knowledge (as adaptation), but 'criticizes' it as well, albeit in a clearly non-Kantian way. The subjective nature of knowledge

[20] Pugh (1978), 109.
[21] Pugh (1978), 110.

does not only result from our limited point of view and from our limited computational power. Foremost, it reflects the fact that knowledge has to send us in particular directions, has to motivate us to behave in certain ways, and has to enable us to make choices between different alternatives. It is to be expected that different organisms, and organisms for example with a different sex or age, will experience the world in different ways, as their experience reflects different options and pitfalls.

This means at the same time, that social knowledge somehow has to deal with these natural differences in interest and perspective. Physicalistic naturalists are often very suspicious about 'folk psychology'—as if it is a completely retarded and prescientific understanding of psychological causality. From the viewpoint of evolutionary epistemology it might be claimed, however, that 'folk psychological' categories may have evolved to refer to the causal factors that really matter. When survival in the real jungle requires adequate physical categories, survival in the *social* jungle could require to some extent adequate psychological categories. Empathy doesn't make much sense when it is based on pure projection. From an evolutionary epistemological point of view it wouldn't be strange to suspect that at least in some mammals, mothers 'understand' their offspring, some dominant individuals 'understand' something about subdominants, males understand something about females, and vice versa. The first person perspective probably not only limits our knowledge about the world, it may also offer unique possibilities to understand other minds.

The evolution of emotions from a more primitive system of punishment and rewards

All in all, consciousness seems to have evolved to motivate us and guide us—even force us—through life. Innate values—in the form of evaluative experiences—are assigned to incoming information in order to enable organisms to weigh their alternatives. What kind of organisms do exhibit such evaluative experiences and in what kind of organisms did consciousness arise?

As we have seen, during large parts of the twentieth century, the very concept of consciousness caused uncomfortable feelings

within the scientific community. Many scientists and even philosophers of mind tended to dismiss the concept, often confusing it with dualistic or supernaturalistic interpretations. There is a difference, however, between first and third person *perspectives* and their metaphysical reification. Progress in neurobiology forces us to accept that consciousness and the working brain are two aspects of the same. Like perspectivism and realism, pluralism of experience and metaphysical monism, go hand in hand.

Obvious as this may sound, it seems to have been the relative difficulty of this idea that resulted in the biggest mistake in the history of psychology: the misunderstanding of conditioning. Behaviorists misunderstood operant conditioning, because they thought of it as a kind of purely mechanistic process which can be understood without psychological concepts: the ambivalence of the term 'reinforcement' did it all. Instead, conditioning and 'reinforcement' seem to be based on the most elementary forms of consciousness: pleasure and pain. Pleasure and pain are probably the most elementary and atavistic forms in which psychological feedback is given to desirable or less desirable behavior. Pleasure and pain can only be rewards or punishments when they are experienced consciously.

One way to understand the origin of consciousness is to observe the ongoing relationship between genes and behavior in those groups of organisms in which behavior became gradually more flexible. At some stage during the evolution of mobile organisms, rigid behavioral programs and systems of reflexes became a disadvantage for longer living organisms. Somehow, more autonomy had to be given to the nervous system in order to enable it to adapt better to unpredictable environments. A whole new level of variation and selection was introduced and organisms started to learn via trial and error. But a feedback mechanism was required to enable the nervous system to know the difference between 'right' and 'wrong'. At this point, rewards and punishment evolved as a kind of biological traffic lights. But a reward doesn't work when it isn't felt and the same goes for punishments. Consciousness evolved as the most elementary framework of biological autonomy: on the one hand, genes still determined the general goals of the organism, on

the other hand, behavior was no longer directly controlled by genes, but could be adapted to a variety of environments.

If all this is true, there is no need for consciousness to evolve at the moment that behavior is completely automatic. In the terminology of Dennett, 'Darwinian creatures' may be completely unconscious. In principle, all kind of sensors do still not require consciousness. Sensors may activate reflex mechanisms or feed neural networks in which particular input-output relationships are broadly defined. Those input-output relationships can be adjusted to a particular environment by adapting the neural connections via trial and error, apparently without the need for consciousness. Even all kinds of smart devices can 'learn' this way. It is unclear to me to what extent the phenomenon of imprinting can be explained in a similar fashion.

In some organisms the flexibility afforded by this kind of learning must have been insufficient. More autonomy was needed. Organisms had to be allowed to learn by trial and error. A more general feedback system was needed in the form of punishment and rewards. Many thinkers have had the intuition that consciousness somehow is dependent on complex feedback loops. This intuition is compatible with the idea that pleasure and pain are part of a feedback loop in which behavior is either reinforced or not. All higher forms of consciousness may go back to this original system.

Humphrey is one of the authors who has reflected on the way in which consciousness could result from a feedback loop[22]. He proposes that consciousness arises from a feedback loop in which the output fibers are gradually projected back into the brain. As a result, he thinks consciousness is dependent on the sensory cortex. Insects and other invertebrates don't have a sensory cortex, and therefore it is not very likely that they are conscious[23]. But Humphrey still doesn't ask himself why conscious sensations would be necessary. Consciousness is more than pure sensitivity: it requires at least two types of sensors: first, external sensors that feed information into the neural network, second, internal sensors that establish the meaning of that information for the organism. The second

[22] Humphrey (1996).
[23] Humphrey (1996), 214.

type of sensor could be named evaluative sensors or interest-sensors. These enable the evaluative experiences which according to Pugh constitute the foundation of both consciousness and values.

Ultimately, conscious experiences arise as a result of the necessity to evaluate the alternative moves an organism can make in reaction to its environment. Successful moves have to be approved, less successful moves have to be disapproved. But this approval doesn't come from the outside: organisms have to feel that some moves are simply not in their own interests while others are. Although, we still do not understand how this approval arises, it probably isn't located on the sensory cortex alone, but is somehow linked to pleasure or reward centers in subcortical regions.

Perhaps consciousness is even more primitive than previously suspected. In recently published research it is suggested that even fruit flies show equivalents of fear[24]. The behavioral patterns of insects are sometimes complex enough to suspect that they are more than pure automata. Of course, it is much harder to deny the existence of consciousness in vertebrates. In my opinion, it is even possible to make a kind of 'emotion tree' in which we can see how behavior has become more flexible during the evolution of birds and mammals (see appendix). The idea is that with each new level of flexibility, new emotions have to be introduced to enable organisms to deal with their new possibilities.

Thus, my proposal is to link consciousness to the emotions. Consciousness enables organisms to weigh options. The emotions form a system of innate values which force animals to do things or to make particular choices as a result of the way they experience particular situations[25]. Emotions work because they are felt, because they force the organism to accept its priorities. Of course, many traditional philosophers have proposed a link between consciousness and reason. Even Dennett falls into this tradition, as he links consciousness to 'mind tools', that is: language. In my opinion, language is a very sophisticated system of sharing information and experiences. It can only have evolved at a point in evolution where there was already subjective information to share. Probably

[24] Rood (2015).
[25] See Johnston (1999).

many animals are conscious, but only humans have art and poetry to share their experiences at a cognitive level and express and describe their emotions and link them to the details of complex ecological and social situations. Reason is—in the famous words of Hume—"slave of the passions": as an ability to calculate, it has, on its own, no motivating power, but when it is informed by the emotions, it enables us to expand our knowledge of the world and to weight our options in it.

Metaphors for consciousness and neuronal workspace theory

To what kind of model does that lead us? Can we build a model about the way consciousness is related to a plurality of unconscious processes? Of course, an organization chart or flow chart would probably be the best way to model all complex relationships within the brain. Such models are being developed by different neurobiologists, but are still highly speculative, and sometimes not very informative. At the moment, the best a philosopher can probably do is sorting out the right metaphors. Such metaphors are referred to by Dennett as 'intuition pumps'[26]: they enable us to use knowledge from one domain to shed light on another.

As a student, a combination of evolutionary epistemology, sociobiology and Pugh's work on the biological origin of human values inspired me to conceive what I pretentiously called the 'dashboard theory of consciousness'[27]. The metaphor of a dashboard seemed to afford an instructive analogy of the relationship between conscious and unconscious. A dashboard provides the driver with a concise overview of the relevant information and of the real options, without offering transparency with respect to the way in which this information is obtained and transferred. Inspired by Pugh and evolutionary epistemology, I stressed that it is not the function of the brain to collect knowledge, but to make the right decisions. It is only logical that the brain would evolve ways to present us information in such a way that it would improve the

[26] Dennett (1984).
[27] Slurink (1998; 2002; 2014).

decision process. Pugh himself also alluded to the adaptive significance of the way in which information is presented:

> To be able to make decisions, the system must learn to *predict* the value consequences of alternatives. The use of distinguishably different values makes it easier to associate specific value components with specific causal factors. One of the functions of rational thought is to classify the motivating drives or values and associate them with the specific types of activity they are intended to motivate. By delivering each drive to the conscious mind in such a way that it is separately distinguishable, the association problem is greatly simplified.[28]

Of course, when I started comparing the way in which we experience the world with a dashboard, most people immediately reacted by pointing out that this idea requires a homunculus behind it. I would then react that the dashboard of experience is curled in such a way that it is both dashboard and the observer, subject and the subjective way in which objects are represented. Mobile organisms are comparable with automobiles, but their drivers are locked up in a globular dashboard, a cockpit, and they are both what is inside and the way in which it is perceived. There is no little guy behind the dashboard, because we ourselves are both the dashboards and the little guy behind the wheel. Above that, the whole concept of autonomy is undermined to some extent when we realize that we ourselves are unable to change the goals that evolution has imposed on us via the design of the dashboard.

Years later, in the nineties, I came across Bernard Baars' book "In the theater of consciousness"[29]. The metaphor of a dashboard and that of a theater have a lot in common: of course, both have the 'problem' of the observer. Apparently, Bernard Baars didn't even bother to defend his metaphor of mind to philosophers like Dennett who rejected the metaphor of a 'Cartesian theater' in his book on consciousness[30]. More important, however, was that Baars came

[28] Pugh (1978), 110.
[29] Baars (1997).
[30] Dennett (1992).

with a somewhat different reason to suspect that consciousness has only limited access to the mind. Baars reasoned that the mind is comprised of a series of specialized information processing units that at some point have to deliver their result to a central information agency. Consciousness is the 'global workspace' in which information from all the senses, from memory and recognition and interpretation centers, comes together and can be subsequently processed.

It is a pity, though, that Baars' model focusses on cognition only. After all, it is behavior that matters above anything else. Baars explains consciousness as a kind of unifying power in which information first converges—'on stage' in his theater metaphor—and subsequently diverges as it is "widely disseminated to members of the audience"[31]. The function of this bottleneck is integration. Consciousness allows us to pick up relevant information from a wide array of specialized circuits that mostly work independently and parallel.

> If we think of the brain as a distributed system with millions of specialized abilities, the question becomes *how to mobilize all of the specialized unconscious networks in pursuit of survival and reproduction.* This is presumably why the unconscious society of the brain requires a stage, a spotlight, and a director. Consciousness, in this view, serves to disseminate a small amount of information to a vast unconscious audience in the brain. It is the publicity organ in the society of mind.[32]

But why is such integration necessary? In his final chapter on "The functions of consciousness" Baars refers to behavior, again and again, but he does not even mention the role of the emotions. Without the emotions, however, it is hard to understand how the brain can set priorities and make adaptive choices. After all, as Pugh shows, without emotions the brain of a rational being would be forced to think continually about the goals of its behavior and about the survival value of each particular decision. The emotions

[31] Baars (1997), 43.
[32] Baars (1997), 45.

constitute evolved heuristic values that enable us to make the choices that matter at each particular stage in our lives. For that reason, I find the metaphor of a dashboard somewhat more informative than the one of a theater, although admittedly both are inadequate and lead to questions about the observer and the person behind the wheel respectively.

The last decade, the theory of Baars has become increasingly popular, and for good reasons. It predicted that consciousness correlates with a large scale integration of activity in diverse regions of the brain, and that is exactly what was demonstrated with the use of fMRI-scanners. The resulting new model was named 'neuronal workspace theory' by the French neuropsychologist Dehaene. Dehaene describes a series of 'signatures of consciousness' with which one can recognize consciousness, even in coma patients. In the future these signatures could even help us to decide which animals are conscious and which ones are not[33].

The recent progress in the study of consciousness was not only a result of new technology, like fMRI-scanning. It was also the result of the realization, by Baars and others, that exactly situations in which visual stimuli do *not* penetrate into consciousness allow experimental psychologists to study the differences with situations in which they do. For example, in binocular rivalry, only one of the two images presented to the eyes, can be consciously seen at the same time. Images which are shown too short, do not become conscious, but do nevertheless have an influence on later observations. In this way, it becomes possible to study the differences between observations that do become conscious and those which do not.

fMRI and older techniques (EEG, MEG) are subsequently used to get a picture of what happens in the brain at the moment that something does become conscious. Consciousness displays itself in a typical sequence in which first different brain regions 'ignite' and then a more global 'avalanche' of information processing starts, finally resulting in a 'brain web' of synchronized neuronal oscillations. In the case of vision, this brain web is characterized by bidirectional causal relations from the visual cortex to

[33] Dehaene (2014).

the frontal lobe, and vice versa, together with a typical pattern of brain waves, both in the lower and higher frequencies.

All this suggests a third metaphor for consciousness: the boardroom. If we compare the multitude of specialized centers in the brain with regular employees in a factory, consciousness is a team meeting of managers in which decisions are made. Most of the time, the workers in different parts of the factory just do their job, but sometimes their tasks have to be coordinated, and the managers are called together in the boardroom. If the decisions made in the boardroom turn out wrong, tensions rise, and new strategies and working plans have to be thought out... Even the directors in the board room don't have the final word: to some extent, they are just puppets controlled by shareholders... (Which brings us back to genes).

Perhaps we can conclude that metaphors can be instructive, but are always inadequate... All three metaphors mentioned suggest that consciousness has a function. Somehow behavior has to be controlled and coordinated in order to adjust it to circumstances. When reflexes and rigid behavioral programs no longer suffice, flexibility is required. Global workspace theory and neuronal workspace theory are plausible models about the way in which coordination is achieved in a brain full of specialized circuits. Although, I personally believe, they still leave out the decisive factor: the emotions.

Affective neuroscience, flexibility and the function of consciousness

Konrad Lorenz once observed that "there are very simple neural processes which are associated with intense experiences and very complex ones, analogue to rational operations, which in spite of that, are beyond experience and even beyond self-observation"[34]. Although unclear about the nature and location of consciousness, Pugh situates the value system of biological decision systems in the midbrain, especially in the hypothalamus.

[34] Lorenz (1973).

> Evidently, the forebrain houses a 'rational' decision system; the midbrain houses a sensory processor and special-purpose computer that drives the value system; and the hindbrain houses an output processor which converts general decision into specific muscle commands. If this perspective is correct, it suggests that the rudimentary value-decision system is a surprisingly old evolutionary invention.[35]

Somewhat later, he argues that "the extreme complexity and refinement of the innate human value system may reflect the operation of ...the... large frontal lobes"[36]. He reasons that during human evolution the midbrain proved too small and that the frontal lobes and sensory cortex took over some of its tasks.

Although modern research has shown that consciousness is by no means exclusively linked to the cortex, few researchers have looked for its origin in deeper parts of the brain. Often consciousness is still largely linked with rationality and neural activity across the cortex, as Dehaene does to some extent. Modern neuroimaging techniques are based on the assumption that the magnetic changes of oxygenated blood reflect all brain activities, but whether all changes in the midbrain are mapped this way remains unclear. Even the much celebrated but sometimes obscure Damasio is not very helpful in creating a truly interdisciplinary model of the brain: his work is mainly based on human brains and he hardly refers to animal behavior[37]. For this reason, the new work of Panksepp & Biven on "The archeology of mind" comes as a revelation. For the first time, the evolutionary anatomy of seven emotional subsystems becomes clear and it is shown that they largely depend on subcortical networks, for example those surrounding the 'periaqueductal gray'[38].

Panksepp and Biven prove their point by pointing out that human babies, "Who are born basically without cerebral hemispheres, can grow up to be affectively vibrant children if they are raised in

[35] Pugh (1978), 134.
[36] Pugh (1978), 135.
[37] Damasio (1999).
[38] Panksepp & Biven (2012).

nurturing and social engaging environments"[39]. Laboratory animals without cortex are even more emotional than normal animals, suggesting that the cortex rather suppresses and corrects the basic emotions than exclusively housing them. The largest part of their book describes the seven emotional subsystems which all are based in the midbrain and do their work both in animals and humans, often showing surprising continuities and analogies in pathological behavior. They show that human neuroticism and psychopathology can sometimes only be understood within the larger framework of the atavistic emotional systems that we share with other animals, which Panksepp & Biven call SEEKING, RAGE, FEAR, LUST, CARE PANIC/GRIEF, PLAY.

Of course, Panksepp's work is largely based on rats. Despite the fact that he discovered that rats 'laugh' when they are tickled, rats probably do not exhibit the full spectrum of social emotions which can be observed in primates. The emotional systems that enable cooperation and culture have probably somehow evolved out of the more atavistic emotions, described by Panksepp: for example, moralistic aggression or indignation probably evolved from RAGE, trust on which all cooperation is based may have evolved from CARE, and so forth. In essence, evolution does seldomly create things *ex nihilo,* as God once was supposed to do, and the same probably goes for emotions. At the same time, it is clear that the emotional systems, which, according to Panksepp, are shared by all mammals, didn't stop evolving in intelligent groups like cetaceans, elephants, apes, and hominids. In some groups of social animals, for example, there are links between group size and the size of particular parts of the brain[40].

For this reason, it should be possible to draw a phylogenetic tree of emotions, which corresponds to the behavioral repertoires of the animals exhibiting them[41]. If I am right in following the intuitions of Lorenz, Pugh and Panksepp and the emotions 'drive' flexible behavior via consciousness, this tree reflects the way in which the requirement of flexible behavior drove psychological com-

[39] Panksepp & Biven (2012), 14.
[40] Dunbar (1993, 2014).
[41] See Appendix.

plexity in vertebrates. Of course, this tree of emotions should not be understood as a hero story, or a story of progress: it is the story of a specialization, a specialization that in one specialized group, the hominids, led to our unique intellectual and moral capacities[42]. It is important to realize that the way in which we express our emotions is phylogenetically contingent. Paul Ekman's classification of the six basic emotions in humans, based on the universal expression of these emotion in humans, have to be seen as one of the many branches of the tree[43]. Of course, any attempt at drawing that tree is, at this moment, a speculative exercise. At the same time, the need for a phylogeny of emotions arises naturally from the realization that even apparently uniquely human properties like morality and cumulative culture have their roots in much older behavioral and cognitive drives.

Conclusion

Consciousness is an evolved property which transforms information into a uniquely perspectivistic, first person experience of the world. In this way, it does not only represent the outside world, but also the unique interests of the individual. The idea that it is simply an epiphenomenon is therefore unlikely: instead, it seems to enable flexible decisions in which a plurality of factors is weighed, as Pugh proposed with his theory of the value driven decision systems. That doesn't mean that all decision processes are conscious: instead, consciousness seems to enable flexible decisions on the basis of a weighing within a *limited overview* of the available information. This can be expressed by a 'dashboard' model of consciousness, which bears similarities to the 'workspace model' of Bernard Baars. The emotions constitute the 'value system' which makes decisions possible: as Lorenz and Pugh suspected and Panksepp proved, they have their basis in older parts of the midbrain, although the elaborate value system underlying the unique cognitive and cooperative abilities of humans also require parts of the forebrain and of the cortex. For this reason, it is possible to

[42] Suddendorf (2013).
[43] Ekman (2003).

draw a speculative 'tree of emotions' which shows their phylogenetic history in some groups of vertebrates: this tree, if designed properly, would illustrate that the emotions underlying cooperation and even morality, are based on an evolutionary older system which is much more widespread in the animal kingdom. This system may have evolved from a more primitive punishment and reward system coupled to conditioning. If this is true, the ability to 'shape' behavior through 'conditioning' and 'reinforcement'—as described by the behaviorists—would in fact be an indication of consciousness—*contrary* to the interpretation of most behaviorists.

Appendix: a speculative tree of emotions

Context	Emotion		Animals
1. Body	Pain ←→ pleasure	Atavistic affects	Fish, Amphibians, Reptiles?
2. Safety → 5	FEAR, FEAR ←→ wellbeing		
2. Food	Hunger ←→ satisfaction DISGUST ←→ appetite		
4. Reproduction → 6	LUST, desire		
5. Exploration, Learning	SEEKING, curiosity, excitement ←→ boredom → SURPRISE	Elementary emotions	Birds? Mammals
6. Breeding care & Attachment → 8, 11	CARE, Love & concern ←→ PANIC, loss		
7. Safety in numbers , social learning → 9	PLAY, sociability ←→ loneliness		
8. Competition for Partners → 10	Jealousy←→ infatuation → love ←→ hatred		
9. Competition for dominance → 10, 11, 12	RAGE, ANGER, pride, self-confidence, arrogance ←→ submissiveness → Admiration		
10. Good & bad luck, winning, losing	HAPPINESS ←→ SADNESS		
11. Cooperation → 12	Trust ←→ distrust, reproach, indignation ←→ guilt, duty, gratitude	Social emotions	Long living intelligent mammals
12. Sympathy + time travel	Empathy, compassion, malicious pleasure, sadism → shame, regret, remorse, guilt		Apes, hominins, Man

A philosopher's speculation about the evolution of emotions, based on behavioral observation and informed guessing. An attempt is made to include both Panksepp's seven emotional systems (Seeking, Lust, Care, Rage, Fear, Panic, Play), based on rats, and Ekman's basic emotions (Anger, Disgust, Fear, Happiness, Sadness, Surprise), largely based on humans. Arrows in the left column (→) refer to hypothetic phylogenetic relationships ('evolves into' or 'enables the evolution of'). Double-headed arrows (←→) in the second column refer to bipolar affects or emotions.

References

Baars, B. J. (1997) *In the Theater of Consciousness. The Workspace of the Mind*, Oxford: Oxford University Press.

Bergson, H. (1908) *L'évolution créatrice*, Paris: F. Alcan.

Bowlby, J. (1980) *Attachment and Loss*, vol. 3, New York: Basic Books.

Chadwick, H. (1980) *Origen: Contra Celsum*, Cambridge: Cambridge University Press.

Churchland, P. (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: The MIT Press.

Damasio, A. (1994) *Descartes' Error — Emotion, Reason and the Human Brain*, New York: Putnam Publishing.

Darwin, C. [1892] (1958) *Charles Darwin, Autobiography*, Mineola: Dover Publications.

Dehaene, S. (2014) *Consciousness and the Brain. Deciphering how the Brain codes our Thoughts*, New York: Viking.

Dennett, D. C. (1984) *Elbow Room: The Varieties of Free Will Worth Wanting*, Oxford: Oxford University Press.

— [1996] (2001) *Kinds of Minds. The Origins of Consciousness*, London: Phoenix.

Dunbar, R. I. M. (1993) "The Co-evolution of Neocortical Size, Group Size and Language in Humans", *Behavioural and Brain Sciences* 16, pp. 681–735.

— (2014) *Human Evolution: A Pelican Introduction*, Harmondsworth: Penguin.

Ekman, P. (2003) *Emotions Revealed*, New York: Henry Holt.

Humphrey, N. K. (1976) "The Social Function of the Intellect", in *Growing Points in Ethology*, edited by P. P. G. Bateson & R. A. Hinde, Cambridge: Cambridge University Press, pp. 303–317.
— (1992) *A History of the Mind. Evolution and the Birth of Consciousness*, New York: Copernicus/Springer.
Johnston, V. S. (1999) *Why we feel. The Science of Human Emotions*, Reading, MA: Perseus Books.
Kant, I. [1781] (1981) *Critik der reinen Vernunft*, Riga: Hartknoch, http://www.deutschestextarchiv.de/book/show/ kant_rvernunft _1781 (last accessed 22 September 2015)
Lorenz, K. [1941] (1983) "Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie" in *Die Evolution des Denkens*, edited by K. Lorenz & F. Wuketits, Zurich and Munich: Piper, pp. 94–125.
— (1973) *Die Rückseite des Spiegels*, Munich: Piper.
Origenes (1985) *Vier Bücher von den Prinzipien*, translated by H. Görgemanns & H. Karpp, Darmstadt: Wissenschaftliche Buchgesellschaft.
Ortega y Gasset, J. [1923] (1963) *De taak van onze tijd*, Den Haag: Leopold.
Panksepp, J. & Biven, L. (2012) *The Archeology of Mind. Neuroevolutionary Origins of Human Emotions*, New York: Norton.
Pugh, G. E. (1978) *The Biological Origin of Human Values*, London & Henley: Routledge & Kegan Paul.
Rood, J. (2015) "Fruit Fly Fear", *The Scientist*, May 15.
Sellars, R. W. (1922) *Evolutionary Naturalism*, Chicago & London: Open Court Publishing Company.
— (1973) *A Naturalistic Theory of Value and Valuation & Agential Causality and Free Will. Neglected Alternatives*, Essays by R. W. Sellars & W. P. Warren, ed. Lewisburg: Bucknell University Press.
Slurink, P. (1998) "Why Consciousness Evolved", in *The Promise of Evolutionary Epistemology*, edited by T. Derksen, Tilburg: Tilburg University Press.
— (2002) *Why some Apes became Humans*, Radboud University Nijmegen, Dissertation. (http://hdl.handle.net/2066/32105).

— (2014) *Aap zoekt zin; waarom wij bewustzijn, vrije wil, cultuur en religie hebben*, Leusden: ISVW Publishers.
Suddendorf, T. (2013) *The Gap. The Science of what separates us from other Animals*, New York: Basic Books.
Vollmer, G. [1975] (1983) *Evolutionäre Erkenntnistheorie*, Stuttgart: Hirzel.
— (1988) *Was können wir wissen?*, Band 1: *Die Natur der Erkenntnis*; vol. 2: *Die Erkenntnis der Natur*, Stuttgart: Hirzel.