

CONSCIOUS MIND AS A LIMIT OF BRAIN-LIKE COMPUTING.

*Włodzisław Duch, Department of Computer Methods, Nicholas Copernicus University,
ul. Grudziądzka 5, 87-100 Toruń, Poland*

E-mail: duch@phys.uni.torun.pl, <http://www.phys.uni.torun.pl/~duch>

Abstract: Fantastic ideas about the mind in general and consciousness in particular abound. Many philosophical efforts are devoted to solutions of non-existing problems. In this paper errors in recent discussions and thought experiments in the philosophy of mind are pointed out. Identical functions imply identical physical construction, therefore one cannot change the brain without changing the mind. Small changes in physical architecture lead to differences in functions. Once these facts are acknowledged hard problem of phenomenal experience becomes a question of finding an approximation to brain-like computing that converges to mind-like behavior. Phenomenal experience arises due to the non-verbal discrimination in the brain-like computing systems and has no special status. Claims of qualia are necessary consequence of brain-like organization of computations, in particular the ability to comment upon physical states of the architecture carrying out these computations. Minimal requirements for an artificial system that should be able to claim qualia are given.

I: Understanding the mind.

What does it mean to understand and what would be a satisfactory understanding of the mind? Recent discussions on consciousness has not paid enough attention to this question. Although the question of truth and meaning has been discussed by many philosophers (cf. Putnam 1975, 1975a) the discussion was concerned with understanding of abstract, intellectual facts, that is quite different than understanding of experiential facts based on the first-hand experience. Understanding of movement was for two thousand years the most challenging problem. Even Kepler still believed in harmony of the spheres and intelligent angels pushing the planets (cf. Gregory 1981). Newton and many of his followers thought that action at a distance is impossible without treating the space itself as an active substance, God's sensorium. Electromagnetic fields without mechanical substrate were hard to imagine, but since we learn about gravitation and electromagnetic waves at school the feeling that we do not understand is largely gone. Do we understand that water is composed from hydrogen and oxygen? It was hard to believe a few hundred years ago but now we do not seem to have problems with it. That multiplication of two negative symbols $(-a) \times (-b)$ gives the same result as $a \times b$ has been pronounced "beyond human understanding" by great mathematicians of the XVII century and yet children at school seem to have no problems with it nowadays.

We do 'understand' such facts now, or at least we have a feeling that we do understand, since our associations are formed during the early school years and our thinking follows these associations in a natural way. The curved space-time concept or the ideas of quantum mechanics are taught only at universities, rather late in the time-course of cognitive development. These concepts are too remote from our life experiences to be experientially understood and thus only a small number of experts enjoy the feeling of understanding them. In case of quantum mechanics, after more than 70 years of discussions, claims of "understanding" are still controversial. There are even more controversial issues: some people claim to understand creation of the Universe in the Big-Bang out of the vacuum (since vacuum is not stable), other people understand creation of the Universe as an act of supernatural omnipotent being. Some people understand that mind is turned off like a bulb when electric processes in the brain stop working, while other people understand that their body is moved by an immaterial spirit. What some people find convincing others find quite irrelevant. Quite a few 'normal' people convert to strange, irrational beliefs, arguing that there is strong evidence where basic facts are missing. The feeling of 'understanding' may be due to rationalization of emotional reactions, filtering new ideas depending on their emotional contents (cf. Ayer 1982 on the role of temperament in philosophy). Understanding requires new facts to fit, at least partially, to existing beliefs. Perhaps the structure of our world view, once created, admits only those new elements that fit to already existing constructions? Sometimes whole substructures of mind (complex theories or beliefs) break or crumble due to the pressure from other substructures that have more weight, but this is a rare process. Are we able to question ideas that were deeply ingrained in our minds at an early age? It took 2000 years to discover some obvious errors of Aristotelian physics and to understand that movement does not require spirits.

To reach comparable understanding of the mind we may have to give up even deeper illusions. Many people are still not satisfied with the idea that mind is a function of the brain. The popularity of simplified models of the

brain (for example left-right or triune brain models, Humpden-Turner 1981) seem to point out to the need for simplicity. Perhaps the brain is too complex and we believe that mind must arise from something more simple and more fundamental (cf. Eccles 1985, Penrose 1994)? Some people find the idea of mind as merely a product of the brain undignified (cf. Eccles 1985), other worry about non-existing problems such as the non-spatial character of thoughts (cf. McGinn 1995, Clarke 1995), proposing a total revolution in science. Philosophy of mind has its share of strange ideas. Although many problems related to mind have been solved by cognitive sciences mind is still perceived as something mysterious, in contrast to physics which has discarded the medieval style of thinking and does not treat movement as something mysterious anymore. On the one hand cognitive, information-processing functions and affective functions, being clearly functions of specific brain structures, are considered to be rather easy to understand. Therefore philosophy of mind has concentrated on more difficult questions, such as consciousness. Is it a problem worth of philosophical considerations or should it be left for neuroscience? Perhaps our feeling of understanding also in this case depends on the concepts we have acquired in the childhood and future generations, taught at school about the brain, behavior and cognitive psychology, will not be so puzzled with the mind being “what the brain does” as some of us are.

A fruitful point of view on the relations between mind and brain is based on investigation of the limits of the brain-like computing. Computational models inspired by neural structures of the brain show some mind-like properties. In the limit of their development these models may produce a conscious-like behavior not only claiming to have qualia, but also believing that they do. What is necessary to create such models and are there any fundamental reasons to think that claims of such systems are different than our own claims that we experience qualia? If mind states are understood as global dynamical states of the brain, states that include influences of the environment (mind is not just in the head, cf. Putnam 1987) as well as the internal dynamics, many philosophical thought experiments leading to paradoxical results have simple solutions. In the next section several misunderstandings in recent discussions on consciousness are pointed out. The third section explains why many thought experiments in the philosophy of mind do not clarify but rather create artificial problems. In the fourth section I will discuss a solution to the problem of phenomenal experience that seems – at least to me – satisfactory and gives me a feeling (no doubt illusory) of “understanding the mind”. The last sections section contains conclusions.

II: Crazy ideas for no reason at all?

The medieval way of looking at the mind, treating it as a kind of substance rather than as a function of the brain, has not disappeared completely yet. Collin McGinn, one of the leading neomysterians, in an article (McGinn 1995) addressing the problem of the non-spatial nature of the mind writes: “... we stare agape in a vacuum of incomprehension”. Why is it that I don't share his feeling? I rather stare at the question “How can the non-spatial arise from spatial?” This very question is based on medieval thinking, so no wonder that the author of the question stares agape. Mind is not a substance like Descartes, following the scholastic tradition, assumed, and the contents of mind is not composed of “things”, *res cogitans*. Mind is a function, a succession of brains states, arising by consecutive patterns of excitations of the brain and body tissue. These patterns of excitations have a certain temporal logic resulting from the physical structure of the brain. The structure exist in form of memory traces and is actualized by the dynamical processes in the brain. One difficulty with understanding of such model is due to the lack of good technical metaphors. Auditory or visual scenes stored on a laser disk are also memory traces (of the camera photocells states) that are actualized by dynamical processes in the electronic equipment, but their temporal structure is fixed, they do not interact with their environment, so the analogy is only partial. Still it makes no sense to describe internal states of an electronic video player as spatial, although physically the system is contained in a small box. These states are defined in a space of electrical activity of various elements, just as the brain states are defined by neuronal activities.

The computational paradigm provides us with an example of non-spatial relations among abstract objects arising during calculations. Putnam (1975a) views the relationship between mental and physical states as analogous to that between logical and structural states of Turing Machines – there is no great mystery here. Spatial concepts are not applicable to states of Turing machines. Unfortunately such machines do not offer good metaphor for brain dynamics, a set of coupled resonators gives a more faithful picture. States of such resonators (one can treat them as models describing activity of neural microcolumns in the neocortex) may follow a complex temporal logic. If these resonators specialize in analysis of visual inputs and are connected to another, linguistic set of resonators that are able to comment on the inner visual and linguistic states, we get a system that describes its states in non-spatial (linguistic) as well as spatial (visual) way. Mental rotation and mental zooming are best understood as operations on visual representations. The ability to read the music score requires acoustic imagery and activates both visual and auditory representations, while solving equations relies on an abstract, non-spatial realm of representations. Influencing brain structures responsible for these acts, for example by electrical stimu-

lation, shows the „structural coherence” (Chalmers 1995) between the mind and the brain states. Brain is localized in space but the contents of mind is not spatial since it is a succession of brain states – not a substance existing somewhere and „viewed from the inner perspective”. Apart from brain states and their temporal flow there is no viewer.

All this seems to be a common knowledge in cognitive science and yet McGinn is not the only one who asks such naive questions. Some physicists and mathematicians also worry about non-locality of the mind. Since mind is non-local and quantum mechanics is non-local perhaps the brain is non-local too (cf. Clarke 1995)? Such reasoning, based on misunderstanding of cognitive science, leads to a grand proposal to “place mind as the key aspect of the universe” or to develop quantum gravity with closed time loops as a basis for a new mind theory (Penrose 1994). The answer to such pseudoquestions will not change our views on such fascinating neuropsychological phenomena as blindsight, unilateral neglect nor thousands of other, well defined problems. The only thing it is designed to explain is the non-locality of the mind, something that from the functionalist point of view does not need any explanation. Strange that no-one asks anymore: how can a sound be transmitted through a wire? You cannot compress a sound or a picture to a wire, it must be a category mistake – 150 years ago it could have been a strong argument, but now everybody understands that it is indeed possible. Technical metaphors help to ‘understand’ (cf. Gregory 1981), but since we do not have good everyday examples of dynamical systems similar to those that provide good models of brain functions some aspects of mind still seem strange.

In contrast to Collin McGinn and a few other philosophers I see no reason to suppose that understanding the mind is beyond our comprehension. The progress in cognitive science is quite rapid, we understand already so many details about the mind, so why should we doubt in our ability to understand more? Of course human mind has a finite capacity and is not able to comprehend all details of even the simplest molecule, once we start to dig deeper and deeper (it is sobering to know that the hydrogen atom is still a hot topic in physics), but it does not mean that the overall scheme of things is not understandable. Asking wrong questions, questions unrelated to experiments, leads to artificial problems and paradoxes that should make us suspicious about the sense of these questions. As John Dewey wrote “intellectual progress usually occurs through sheer abandonment of questions together with both alternatives they assume ... We do not solve them, we get over them”.

In recent years discussions in philosophy of mind focused on qualia, or the problem of qualitative character of phenomenal experience. This is the hard problem that cognitive science has not addressed, and since it is impossible to imagine consciousness without qualia it became a most debated topic. Chalmers (1995,1996a) has not only identified the problem but also proposed a solution based on the hypothetical dual aspect of information. He has formulated two principles which are in agreement with cognitive science. The principle of structural coherence says that the structure of conscious experience is based on the contents accessible to awareness. This is what most cognitive scientists subscribe to. The principle of organizational invariance is basically the functionalist credo: what matters are the specific patterns of causal interactions between the components, not the substance the components are made from.

After defining the problem and the two principles Chalmers proposes a non-reductive solution based on the alleged physical and phenomenal aspects of information (Spinoza formulated similar dual aspect idea in 1677). According to this idea physical processing of information brings along the phenomenal aspect as well – a naturalistic dualism, one may say. Why do we need to add this idea to functionalist approach and does it have any consequences? Special states of the brain are still necessary to explain the difference between conscious and unconscious information processing. The brain processes almost all information in an unconscious way – actually in situations demanding fast reactions and intensive information processing consciousness seems to be largely absent, while during contemplation of simple objects or ideas, when much less information is processed, conscious feelings are the strongest. Dual aspect of information is an unnecessary extra addition to the functional understanding of the mind, an addition that has no consequences and introduces complications related to the very concept of information. It does not explain anything, but rather forsakes explanations claiming in essence that phenomenal experience exists because God wanted so. Yet quite a few people took this suggestion seriously, starting to call it a “theory” (cf. the discussion following Chalmers 1995 paper, summarized in Shear 1997, Chalmers 1997, for detailed explanation see Chalmers 1996a).

A common problem with many solutions to the “problem of consciousness” is that they never address the important questions such as: what type of systems seem to have minds? What architecture is necessary to create systems that will claim to have qualia? What are the real problems in cognitive science, besides the phenomenal experience, that such systems may solve? Is it relevant to color vision research, or maybe understanding of schizophrenia? Is it possible that we have overlooked non-local quantum effects in the brain on such a grand scale that the mind is produced? Where is the dual aspect of information gone when the brain is under anesthesia

and no experience whatsoever is produced although information processing goes on? Philosophical discussions frequently stray away from real problems because too much faith is put into thought experiments that cannot be realized even in principle. Once we agree to follow the reasoning based on such thought experiments one is stuck with artificial problems. Therefore in the next section a critical analysis of some thought experiments is made.

III: Thought experiments and what's wrong with them.

The principle of organizational invariance is invoked so often in philosophical discussion that it may be useful to point out that it is basically empty. The number of different chemical elements at our disposal is rather limited and no other element can interact in the same way as carbon atoms do. Therefore the functionalist critique of carbon chauvinism is not valid: carbon cannot be replaced by other elements without changing the system in a fundamental way. What does it mean that the functional structures are identical or that their patterns of interactions are identical? Silicon structures and carbon structures are quite different, their chemistry is different, atomic interactions are different, binding to molecules is different. For example, there is no way to create neural receptors or propagate neurotransmitters in silicon in exactly the same way as in biological structures. The problem here is the accuracy of approximation that may be achieved: only very rough features of carbon-based patterns of interactions may be represented in other type of structures, whatever they might be.

This does not mean that intelligence or some form of mind cannot arise due to the interactions of complex forms of matter based on silicon or other non-carbon compounds. However, such minds cannot be identical with our mind, since the problems of accurate approximation of our brain functions is very difficult. A change in the level of a single neurotransmitter in our brain is sufficient to change dramatically contents of our mind including qualia of conscious perception. If in a complex system small structural changes result in large changes of its functions, replacing parts of such system by something 'functionally equivalent' may not be possible. Accurate approximation of functions of such system by a different system may be hopeless. Thought experiments in which small neural circuits are replaced by silicon circuits step by step, until the whole brain is converted to silicon, have no sense. Even small disturbance of the integrity of the brain has a profound effect on the phenomenal experience.

But what if the replacement had identical functional properties? – one may still insist. Replacing neurons or the biochemical substances of the brain with silicon elements or anything else of this sort has to change their quantum states and thus interactions with neurochemicals. The modified system will not function in exactly the same way. The only possible replacement with identical functional properties must contain identical molecules in identical quantum states. Arguments based on reasoning that "in principle" it is possible should not be treated seriously. In this sense quantum mechanics is indeed fundamental to the brain/mind understanding. Although I do not believe that coherent macroscopic quantum states are needed to understand consciousness it is obvious that interactions of neurochemicals determine the state of the brain, including neural patterns of activities. Silicon or other artificial brains may only be rough approximations to what the biological brain does, but there is no reason why silicon-based brain models of growing complexity should converge to the real thing. Different brain structures lead to different minds. To understand better the hard problem of consciousness we should investigate what type of artificial brains are capable of beliefs that they have qualia.

Replacing carbon with silicon or some other stuff is not the only thought experiment that is fundamentally wrong. One of the favorite arguments of philosophers is based on the **inverted qualia**, or at least on their possibility. Let us replace the brain wiring in such a way that connections with green and red cones are reversed and all reactive dispositions are also reversed. Then the experience of green and red is reversed – a case of inverted qualia. Such thought experiment is possible only if the brain is understood as a kind of mechanical device. There is no reason to suppose that mind contents, identical with the global state of the brain, stays unchanged after such manipulations. If all memory associations to the green and red color are removed will the experience of green and red still differ? What makes qualia different? In case of colors many factors are important: retinal signals, activation of V4 and other visual areas, coupling of the visual cortex with the limbic system. All these and many other components of the global brain states create differences in mental contents. There is no way to separate phenomenal experience from information processing. Any change in local information processing must influence global brain states. Either dynamical states of the brain are identical and thus the experiences are identical or they are changed and thus experiences are changed.

Is there any reason to believe that pure experience of qualia, independent of brain states, may exist? Thought experiments with inverted qualia or absent qualia (cf. Chalmers 1996a, Hardcastle 1996) – zombies – are based on fundamental misunderstanding. There is no way to manipulate the brain tissue replacing small circuits here and there without changing the dynamical states of the brain – the state of these circuits themselves is a part of

the global brain state. Our phenomenal experience does not result from excitations of pineal gland or higher areas such as frontal lobes, but from the state of the whole brain. Replacing a small part of the brain, changing sensory information, or even making changes in familiar environment, must change brain/mind states. A damage to the V4 visual area not only impairs information processing related to color but may disable experiences of color in dreams or imagery (Sacks 1995). It takes time to learn the meaning of modified brain states: the limbic structures must assign them new values and the neocortex must assign them new cognitive categories. Philosophers try to go around this problem proposing a series of small, gradual changes, replacing one neuron after another. Since such small changes should not influence the mind states perceptibly one may think that the whole procedure is plausible. Can we guarantee that a large number of small changes does not lead to significant change in qualia? How can I be sure that my qualia now are the same as they were 30 years ago?

The “small changes” argument has been used in many different contexts. For example, Putnam (in an appendix to his Royce Lectures 1998) writes that the statement: “if two phenomenal states seem to be the same than they are the same internal phenomenal state” cannot be true, because indistinguishability in appearance is not a transitive relation while being in the same state is. The same argument has been used earlier by Ayers (1982). If an experiment is performed with almost continuous stimuli we may not be able to distinguish two cards that have almost the same color, so cards 1 and 2 will be judged of the same color, and cards 2/3, 3/4 etc. also, but cards 1 and 4 will already be distinguished as of different hue. Putnam draws a far reaching conclusions from this argument: qualia cannot exist, there is no “highest common factor” in identical experiences. This type of argument is in fact a modern version of the Zeno paradox. Although the convergence of infinite series to a finite limit is now well understood philosophers still find it difficult to deal with continuous or infinitely divisible processes since our minds offer only a finite resolution of the incoming stimuli. Conclusion that Putnam draws from the experiment with colored cards is indeed paradoxical. If the brain states for cards $2n-1$ and $2n$ are identical for all $n=1,2,..50$, than brain states for card 1 (which is completely white) and card 100 (which is completely red) must be identical. We know that this is not true, since different areas of visual cortex (in particular in the V4 area) are active in both cases. Perhaps we may even measure a small difference in brain’s activity for cards 1 and 2, although they seem identical.

Formation of categories and the ability to discriminate is an active area in cognitive psychology. In a few cases we have results of experiments done with monkeys, showing what happens to the neurons during categorization processes, in other cases we may rely on brain imaging techniques, so philosophical arguments may be confronted directly with neurophysiological data. Brain dynamics is in some cases so fine-grained that it should be treated as continuous. On the other hand our ability to distinguish between different brain states is coarse-grained. Identical phenomenal states imply similar brain states, but brain states are never absolutely identical – what we are able to report, also in the internal, “feeling” sense, is coarse or at best medium-grained. Global dynamics, responsible for the contents of the mind, is much simpler than the total dynamics of the brain/body states – on any two occasions many brain substructures may be in quite different states despite the fact that the global dynamics and thus phenomenal experiences are quite similar. Usually everything that belongs to an attractor basin of one category is labeled by one symbol – such as “chair, face, color” – in a coarse-grain fashion. Since the number of symbols used for communication is lower than the number of distinguishable brain states these states must be discerned using an additional, non-verbal internal labels, “qualities” of experience. These non-verbal labels are medium-grained: one phenomenal state corresponds to a certain range of global dynamical states of the brain. Comparing phenomenal states we cannot use classical logic claiming that they are identical, we can only determine a degree of similarity and thus in fact use fuzzy logic (modal logic may be used in some situations). The perceived degree of similarity may be related to the transition probabilities between corresponding attractors in the brain dynamics (Duch 1997).

Classical logic and discrete symbols are not a good way to approximate continuous brain states. It is difficult to speak using symbolic language about essentially continuous processes. If linguistic symbols have to be used (differential equations offer an alternative here) fuzzy sets and fuzzy logic are much better than sharp divisions. “The same” in case of Putnam’s example means simply “sufficiently similar to be indistinguishable by a typical person”. Similarity relation does not need to be transitive. Estimation of the “same height” or more precisely “similar height” for human body has an accuracy of about ± 1 cm, but such similarity relations are definitely not transitive (fuzzy memberships take this into account in quite natural way). Such definition of “the same” extends to comparisons of brain states of different people and of different species. Human pain is similar to monkey pain, but much less similar to octopus pain, and so is the physical structure. The degree of similarity is at some point sufficiently small to use different word, so we use words like “mouth, snout, beak”, or other names to point at the body part with similar functional properties. Why then should one universal and identical physical state corresponding to such concepts as pain exist? We should be satisfied with concepts such as “human-pain” or “mollusk-pain”. There is no sense asking if mollusk-pain is real, but we may determine how similar it is to hu-

man pain looking at the behavior and biochemical processes in humans and mollusks. On the other hand it would be naive to hope that two people experiencing similar phenomenal states (or even the same person at different times) have similar brain states – this is not true even in the case of sniffing rabbits (Freeman 1996). What is important are relations between different internal representations (second order similarities) rather than the actual form of these representations.

What is it like to be someone else? There seems to be a deep-rooted misconception when we try to answer this question. There are at least two kinds of understanding, intellectual and experiential. Intellectual understanding, involving mostly frontal and temporal lobes, allows us to create models of the world and communicate with each other on that basis. Experiential understanding, engaging mostly the limbic system, allows us to share the feelings of our children and our friends. Experiential understanding between two minds requires certain ‘resonance’ between brain states responsible for the contents of these minds. The necessary condition for such ‘resonance’ is defined as follows: dynamics of both brains should admit attractors with similar relational structure in such a way that an approximate correspondence between potential states that the dynamics of each brain may assume could be established. The two brain states do not have to be similar but the structure of transitions between different neurodynamical states should be roughly similar. If this is true than the resonance may occur, and the experiential understanding established, when both brains are in dynamical states that in the network of their relational states correspond to each other. We are able to share such mind states to a high degree with our family members, with other members of the same culture, to a somehow lesser degree with members of different cultures and to even lesser degree with animals, since not only their minds are formed by very different environment but their brains and their senses are physically different.

There is something it is like to be a bat and something it is like to be a man, since “to be” means to be a series of mind states produced by the brain of a bat or of a man, implying a subjective view. Intellectual understanding requires an objective, external description and one is not reducible to the other. To know what it is like to be a bat for a bat requires a bat’s brain in place of our own. Even though it is impossible we may form a fairly detailed description of bat’s internal states and achieve intellectual understanding in modeling bat’s behavior. When we find a particular state of the brain we may infer that a particular experience, whatever that might be for a bat, is correlated with it. Since humans share several needs with bats, such as the need for food and sleep, drawing on our own experiences we may assign reasonable interpretations to the behavior of bats. There is no deep mystery in the celebrated question of Thomas Nagel (1974) “how is it like to be a bat”. Nagel himself admits that perhaps all robots complex enough to behave like a person, would have experiences. His may objection is not to the physicalism itself, but rather to the lack of “the beginnings of a conception of how it might be true”. This is precisely what I try to provide in the next section.

Another famous thought experiment concerns **Mary, the colorblind neuroscientist**, who gains color vision and learns about red (Jackson 1982). There are inner facts that are over and above the physical facts, but the conclusion that physicalism is false because knowing everything about neuroscience does not imply knowledge about qualia, is premature. Dennett’s solution (Dennett 1996) is to deny the problem by claiming that to know everything means to be able to correlate the qualia with brain states. In his version of the experiment Mary is presented with bright blue banana and immediately recognizes the fact (perhaps with access to the maps of activity of the V4 visual area it could be done even today). Dennett concludes that the story does not prove that Mary has learned anything new. She has not learned anything new only in the sense of verbal, intellectual learning, but certainly her brain, stimulated for the first time by a color light, assumed new dynamical states, so she must have felt it as a new experience. Her previous knowledge was abstract, symbolic, engaging temporal and frontal lobes only, not occipital cortex. There is no great mystery in the fact that new brain states are experienced as mind events having new qualities. People that were born blind and gain their sight after they are grown-up certainly learn quite a lot, and it helps them little if they have great intellectual knowledge of geometry. Inner facts are real, although they are only shadows of neurodynamics.

The **Chinese Room experiment** of John Searle (1980) was designed to show that mere computations are not sufficient to bring real understanding. Since a person locked in the room and shuffling symbols to correlate incoming Chinese signs with the outgoing Chinese signs does not understand neither questions nor answers, therefore formal systems based on rules and symbols are incapable of real understanding. Could such a person find anything in our brain, turning into a demon observing neural processes? Already Leibnitz in his *Monadology* understood that it is impossible. Leibnitz asks us to enter the thinking, feeling and perceiving machine just to find there mechanisms rather than floating thoughts. We can understand (in experiential sense) only the systems that have minds of similar structure to ours, by ‘resonating’ with such minds, or trying to assume similar dynamical states of our brain structures. One way of creating such resonance state is through language and observation of behavior. The Chinese Room experiment does not try to discover the mind of a system by bringing our

mind in resonance with it, and thus it does not teach us anything about the mind of artificial system. In this sense it is not a good test for the mind: it fails humans as well as machines.

The Chinese interlocutor may have an impression that there is an understanding mind in the room, but unless the whole system is really able to resonate with their minds this impression will break down sooner or later. It is very easy for humans to lose understanding if their brain processes are not properly synchronized, if the dynamical states with all the background references do not arise. There is no way to represent accurately the states of a dynamical system by rules, therefore any approximation to understanding based on experts systems shuffling symbols will have a serious problem with convergence. If I learn the sound of a single Chinese word, even without learning its meaning, by observing other people and behaving as they do I may slowly acquire at least partially correct behavioral association and start to understand the word in an experiential way. My response or my understanding will be nothing more or less than the state of my brain and the rest of my body. Rules and computations are not a replacement for real physical states of brain/body.

If thought experiments, such as described above, are rejected, no serious objections against naturalistic solution are left. Mind is a function of the brain and the brain-like computing may approximate many mind-like features. An artificial dynamical system capable of assuming brain-like states could resonate with our mind and could achieve real understanding, but construction of such a system is much more difficult than construction of rule-based expert system. Is it possible in the limit of more and more faithful brain models to reach the level at which these systems will claim to have qualia?

IV: Approximating mind and converging to consciousness.

A fruitful point of view at the problems in philosophy of mind is to start with the question: what kind of systems seem to have minds? The answer is rather obvious: mind functions and brain complexity are closely connected. Sophisticated behavior requires sophisticated brain. If panpsychism was true psychopharmacology could not work and people in coma or under deep anesthesia should be conscious. Active brains are the only systems that are undoubtedly associated with minds. How then should we understand brain functions? Understanding always requires simple models but oversimplification or wrong metaphors leads to insurmountable problems, as I have tried to show in the previous sections. Computational, Turing machine metaphors are not quite applicable to the brain dynamics but we have little experience with multidimensional dynamical systems that should provide better metaphors.

A good way to look at the problem of understanding the brain dynamics is to treat it as an approximation problem (Duch 1997). Several levels of approximation are summarized below, starting from the highest, or the most crude level to approximate brain dynamics. In many experiments specific questions may be answered and understood at one of the higher levels (sometimes only at highest levels), although in some situations lower level processes may “shine through”, pointing to the need of a more detailed models.

Long-term behavior. At the highest level of approximation behavior may be understood in symbolic terms, as a set of actions modeled using rule-based system or a finite automata. According to Newell (1990) this is an appropriate level for cognitive science. At this level cognitive models approximate knowledge-based systems using symbolic processing, solving problems requiring reasoning, thinking, and (rarely) creativity. This is the level of rational mind that artificial intelligence tries to model. Search processes in problem spaces are the most common method of AI reasoning and there is some evidence that similar processes are indeed used by humans (Newell 1990).

Computational power (speed as well as memory) of the brain is still on a factor of 10^4 or more larger than that provided by the most powerful supercomputers today, but even this meager computer speed and memory is sufficient to beat humans in such intellectual pursuits as playing chess or theorem proving (cf. computer generated proof of the Robinson's conjecture after 60 years of human unsuccessful attempts, McCune 1997). Algorithmic processes seem to have sufficient power to explain abstract thinking (cf. Penrose 1994 claims to the contrary and Putnam's 1995 rebuttal) and there is little doubt that increasing computer power and increasing complexity of knowledge bases will lead to systems that will exceed human level in many more fields than games or some branches of mathematics. Creativity at this level is rather limited since unexpected associations are difficult to form when symbolic representations are used. Only limited understanding of natural language texts has been achieved. It is not quite clear what processes in the brain are being modeled at this level – presumably it is the long-time dynamics of the brain (more than 10 seconds), especially of the frontal and temporal lobes and limbic

structures, taking place during thinking processes. Baars (1993) blackboard architecture of cognitive processes provides a plausible model here.

Short-term behavior. Associations and immediate (“intuitive”) behavioral reactions (on the order of one second) determine contents of mind. Dynamics of large brain structures is involved here. A crude approximation to this dynamics is offered by probabilistic and finite state automata, while more detailed approximations require large-scale models of transcortical neural assemblies and computational maps in the brain. Many experiments in cognitive psychology, involving associations, similarity judgement, object recognition, categorization, operate at this time-scale. Psychological spaces offer a natural arena to model such processes and to talk about mind events. Although precise internal features of cognitive representations are not known they certainly are provided by brain modules such as sensory maps and various specialized areas, analyzing the incoming information and providing sensomotoric responses. Mental events modeled using psychological spaces (P-spaces) are merely shadows of neurophysiological states of the brain (Duch 1997).

Associations between attractor states of brain dynamics gives structural properties to experience. Internal representations in P-spaces should not be modeled as iconic pictures, sounds or thoughts but rather probability density functions determining probability of combination of particular microfeatures of internal representation. Areas with high probability values correspond to objects that are well known and actions that are frequently repeated. The ‘state of mind’ in this model is just a pointer to a combination of states of specific brain substructures, allowing to recreate the neurophysiological state of the brain and body. The topographical structure of the P-space determines the temporal evolution of the mind state and thus the brain/body states. One can introduce precise language to speak about mind states, connecting psychology and neuroscience (Duch 1994-97). Similar concepts and beliefs have similar topographical representation in the P-space, corresponding to similar structure of neurodynamical attractors, and resulting in similarity judgements in psychological experiments. Using very crude approximations to brain states, approximations based on finite state automata or Bayesian belief networks, we should find similarity of relations between concepts and beliefs, although brain states of two people entertaining the same concept may be quite different. The actual form of internal representation has nothing to do with the represented object as long as the second order similarity relations are preserved, i.e. similarities between representations should approximate similarities between objects (perhaps “representation” is not a good word here, cf. Freeman 1996). For example, using multidimensional scaling technique to project high-dimensional feature vectors into two dimensions we (Duch and Naud, unpublished) have recreated experimental semantic distance charts for animals obtained by Rips, Shoben and Smith (1973). Such relations may be hard to find at more detailed levels of modeling.

Lower, more detailed levels of approximation of the brain functions are needed to understand such processes as the dynamics of learning, moods, formation of habits and addictions, some memory effects, influence of various drugs on brain functions, as well as neurophysiological responses to cognitive stimuli. Models explaining these effects operate on short time scale and are based on small recurrent spiking or graded response neural networks, or on simplified dynamical systems. Understanding of conditioning and basic forms of learning (such as LTP and LTD) requires detailed models of single neurons, electric compartmental models of these neurons, understanding of molecular and synaptic level, biophysics of ion channels and membrane properties, down to the level of genes and their influence on neurochemistry. Unfortunately all these details show up sometimes at higher levels making the higher level approximations brittle. For example sudden stress situation may dramatically increase the amount of neurotransmitters distributed around the brain from several brain stem nuclei, influencing the activity of single neurons, small groups of neurons (microcolumns), forming in turn whole subsystems and finally create changes in the topography of the P-space, corresponding to changes in a global dynamics of the brain's electrochemical processes. Is it possible that another physical realization will give rise to identical dynamical states with the same functional relations? The more complex the system and its functions the more difficult it is to implement it in an alternative way.

Only very rough models of human sensory systems have been made and models of various types of human memory are still quite primitive but the more we know about the brain the more accurate models are made. This modeling process seems to converge slowly to a detailed functional approximation of many brain functions. For example, most visual illusions are well understood, and inspirations based on understanding of color vision and object recognition already lead to practical applications. Even simple models of distributed connectionist memory exhibit a number of properties (such as contents addressability, noise resistance or independence of retrieval times from the number of items stored) that makes them more like human rather than computer memory (cf. McClelland and Rumelhart 1987, Ruppin 1995). Better approximations of brain functions should lead to better approximations of mind functions. Understanding of neuropsychological syndromes is possible on the level of information processing by different brain modules. Even Capgras syndrome, in which sufferer is absolutely

convinced that a family member, a friend, an item of personal value or one's own self has been replaced by alien imposter, is understandable. Cognitive (neocortical) and affective (limbic system) functions become dissociated and despite a perfect recognition there is no emotional feeling associated with the person or object that is recognized. Psychiatric problems require either understanding of the effects of neurotransmitters on the overall brain dynamics (in depression) or understanding of synchronization of different brain areas (in schizophrenia). Hallucinations are understood as breakdown of associative memory and come quite naturally in computer associative memory models (Ruppin 1995).

More detailed approximations to the models of the brain will certainly lead to more precise explanations of all features of our behavior and inner states of mind. Brain-like computing leads to mind-like behavior – the only remaining problem being the problem of phenomenal experience. I have already mentioned how hopeless this problem seems to be to many experts. Proposals to abandon all science and start anew have a curious appeal to many intelligent people (Gardner 1996). Other people do not see any problem at all, hoping that the qualia problem will somehow vanish in the same way as the problem of life has already vanished. Dennett (1996) writes about ‘cuteness’ as another example of artificial hard problem. However, such concepts as ‘life’ or ‘cuteness’ are abstract and qualia are experiential, independent of any intellectual concepts (cf. Chalmers 1997). Animals and small children, even before they are able to form concepts, seem to be able to experience qualia. Therefore the problem should not be dismissed so lightly.

To claim that the qualia problem is just a matter of dispositions, as Dennett does (1991), misses the point. A simple computer program may be disposed to say ‘ouch’ whenever the computer screen is touched but I do not suppose that the program or the computer feels anything. Such solution may be satisfactory to behaviorists but not to cognitive scientists. Perhaps we should ask what level of approximation to the brain functions is sufficient to create systems that will claim to have phenomenal experiences involving qualia, and believe that they really have them. The brain is composed of tightly coupled neural assemblies forming neocortex, cerebellum and other substructures (such as subcortical nuclei). The brain state cannot be defined without ceaseless input from the memory and memory-based reactive dispositions. Even if we are not aware of all the background context forming the brain state (if information about this context is not sufficiently salient) it is still there.

A **necessary** condition to claim that we are conscious of something is the ability to report this fact in any form (not necessarily verbally) to oneself or to the external world, even if this ability lasts for a brief moment. ‘Reporting to oneself’ means that being conscious of something has an influence on the highest control level of the system, i.e. it may directly influence behavior. Although a large number of processes goes on in my brain very few have direct influence on its global dynamics. Processes that cannot ‘be reported’ (i.e. have no influence on the global dynamics) externally or internally to oneself are not experienced and may have only indirect influence on behavior. Such processes are not accompanied by any qualia as long as they do not report anything at the global level. This morning I felt surprised hearing myself using a rare expression. This experience brought along interesting qualia – it could have been an internal experience, a thought, if the speech sounds were not made audibly but were propagated only internally. In any case a representation of a word has been activated in my brain and influenced my behavior – I could report this fact to myself as well as externally. Perhaps many other rare words were activated in my brain on the same morning but were not experienced – were not reported externally nor internally. I can see a red color and associate a particular feeling with its hue. This is not just a matter of dispositions, but a real process in which the neurophysiological state of my visual cortex, my associative cortex and limbic system, all contribute to the global state of the brain. I can report this fact to myself. Whatever dominates the global dynamics of the brain appears as the contents of mind. This is the highest level of control and the information available at this level has an influence on the subsequent brain states, leading to verbal comments and non-verbal states.

Since the global dynamics of the brain constantly evolves, focusing attention on some sensory or imaginary stimuli requires constant non-verbal updating of this dynamics to maintain the experience for some period of time. Watching red object the outputs from visual system moment after moment contribute to the global dynamics of the brain and are mixed or ‘tinted’ by background processes. Because of the tight coupling of all these processes in the brain, even the states of a low-level visual cortex (V1) and the intermediate nuclei (LGN) are modified by higher level processes. The up-going and the down-going streams of information self-organize to form the dynamical state of the brain (Ullman 1996). Primary consciousness or phenomenal experience of the sensory stimuli results from this constant updating of the global dynamics. Qualia result from the ability to discriminate between different states of the brain, states that differ not only by the information coming from the primary sensory areas but also include the background processes. Although ‘reactive dispositions’, as Dennett (1991) calls it, are of primary importance, qualia result from discrimination of real brain states. Dispositions or activation of memory traces influence the global dynamics of the brain, and thus the mind state. These discrimi-

nations probably take place in subiculum (Gray 1995) and influence the global dynamics of the brain. Qualia do have functional role: recognizing this particular color as the one that gives me a particular pleasure to look at I make some decisions. Recognizing sweet or bitter taste I eat or spit. The spitting reaction is frequently automatic, but the memory of bitter taste, allowing to avoid unpleasant experiences later, is kept. But why does it feel like something instead of being processed automatically, “in the dark”? Why do qualia exist?

One reason for existence of qualia is due to the disproportion of the small working memory based on the dynamical, short-lived states of the brain dynamics, and the large-scale unconscious information processing. The taste information has to be compared with all the memory traces in our brain in fraction of a second. Working memory implemented by dynamical processes in the brain is used to broadcast this information (Baars 1988) around the brain, activating similar memory traces in the neocortex and using associations stored there. Discrimination of information into different types makes the comparison easier. Nonverbal labeling helps in this process by assigning brain states a “feeling” interpretation, allowing to estimate the quality of the information. Qualia are non-verbal labels associated with information accessible at the highest control level, arising through persistent updating of the global dynamical state of the brain with information from selected brain and body substructures. We have different feelings because mind state, or the global brain dynamics, is modified in a different way by stimuli of different sensory and internal modalities. Since these modifications are essentially continuous and the number of discrete symbols we are able to use is limited there must be a way to interpret or “feel” the differences. If I sit on a cold stone gradually I feel colder and colder. To be able to reason and take action I need a mental representation of this continuous change, a mental representation of the feeling of getting colder. At the highest control level of the brain’s dynamics this feeling is experienced – I can comment on it verbally but also notice it on non-verbal, qualia level. Rats discriminate different tastes very precisely – to remember and compare tastes they have to label them and it has to be done non-verbally. Thus they must have “a feeling” for different tastes. Why do we lose taste when the sense of smell is blocked? The taste buds provide all the information, the brain processes it but the qualia are gone. Smell is the necessary background for the taste in the global dynamics.

Continuous nonverbal internal reporting is inseparable from attentional mechanisms. Even if the information about the taste of an ice cream reaches the brain if it does not influence the global dynamics, because attention is directed at thoughts or visual experiences, we suddenly notice that the ice cream is gone but our experience of its taste was gone much earlier – in fact it is very difficult to experience something like a taste for a longer time. Most people do not even remember how their last meal tasted. “Directing attention” is just a figure of speech since in reality this is a part of global dynamics, there is no homunculus to “direct”. There *is something it is like to be* in different states of mind, since states of mind are real physical brain states on which the brain may comment. The only way to learn how is it to be something else is not to pay attention to information processing by neurons or computational modules – it is necessary to ‘resonate’ or be in similar states with the other system.

Our general feeling of personal identity seems to be strongly coupled with proprioception. Proprioceptive information is constantly provided by the spindles, one of the two types of the skeletal muscles. The brain has developed in order to control the movements of the body and to analyze the sensory feedback information about the outcomes of movements. Cotterill (1996) discusses the merits of anticipation of such outcomes and the need for an internal model of the body to solve ‘what if’ problems of motor control. Such model was an important step towards construction of the self. Even simplest body acts require massive information processing. Simulations of abstract reasoning, for example in chess or mathematics, is more advanced than simulation of movement control for autonomous robots – construction of artificial rat is still rather far off.

Can we separate ‘pure experience’ from the real, total experience? And if not, how can one claim that phenomenal experience has no functional role? Pain and pleasure have certainly functional roles and perception of colors has of course an evolutionary advantage. How could we discriminate colors if we had no qualia associated with them? Neurophysiological experiments in which the stream of information going from higher to lower visual areas is blocked are possible on monkeys and the results indicate that the lower areas (even the primary visual cortex) are not able to form the same activation as in the normal case. Although there are no experiments with humans one may guess that without help from memory and higher information processing areas qualia will be quite different, more subdued or even perhaps vanishing together with the vanishing sensory experience. An example of vanishing qualia accompanied by greatly reduced behavioral competence is found in blindsight – information processing is greatly impaired, global brain states are drastically changed. This is just what one should expect if mind contents is identical with the contents of the global dynamical state of the brain. Although some information from the eyes reaches the brain and is used to select responses it has no influence on the global brain dynamics, the responses are selected subconsciously. The residual visual information is not discriminated against other types of information since in the normal brain there was no reason to do it – hence there are no

visual experiences. There are no known cases in which qualia have disappeared but the behavioral competence stayed unchanged.

Qualia result from discrimination of continuous brain states, they do not have any special status among other information processing states. Necessary minimal conditions for brain-like computational systems – let's call it an **artilect** (artificial intellect) – that will claim to have qualia include:

- 1) Working Memory (WM), a recurrent dynamic model of current global brain state, containing enough information to re-instate the dynamical states of all the subsystems;
- 2) permanent memory for storing WM states;
- 3) ability to discriminate between continuously changing states of WM;
- 4) mechanism for activation of associations stored in permanent memory and updating WM states;
- 5) ability to report on the actual state of WM;
- 6) representation of 'the self'.

Such system would be able to comment on the actual model brain state and describe it using expressions such as "feels like sweet candy" or "its a very unpleasant red color" because of its similarity to the color of blood. It could insist that it has an experience, because the experience will be a real dynamical state of its WM. Artilects which will be complex enough to have some internal representation of 'self' may be absolutely convinced that they feel qualities of their dynamical WM states and will report it, not only verbally to the outside world, but also nonverbally, using qualia to reason in a non-symbolic way. Claims of qualia are necessary consequence of brain-like organization of computations, in particular the ability to comment on the physical states of the hardware carrying these computations.

Increasing the faithfulness of color or sound reproduction does not lead to perceptible differences at some level of representation. Is it possible to reach such accuracy of approximation that to our mind (with its characteristic final resolution of information) artilects will be indistinguishable from real minds? How would we know that such artilect has no phenomenal experiences despite its claims to the contrary? It may not be able to answer Gödelian questions related to its design (Penrose 1994) for the same reason that we are not able to: our memory is too small to store the detailed specification of our brain. The working memory of artilect may have quite small capacity, just like our working memory. Perhaps to have real qualia we should have real physical processes creating and sustaining the WM states. An analog implementation using electrical processes could be used, but perhaps silicon processor implementing WM functions would be sufficient. Artilects have to accept the states of their mind as real, although for us their minds may look just like virtual reality, without anyone inside. Is the difference between our phenomenal reality and virtual reality only a matter of the quality of approximation? Temporal and spatial resolution abilities of our conscious experiences do not allow us to have an insight into the neurodynamics behind the scenes. We may deceive ourselves thinking that behind our qualia there is something more than nonverbal reporting, invoking of associations and discrimination of continuous brain states.

V: Conclusions.

In this paper I have argued that:

1. Similar functions may be realized by different systems with similar organization of causal interactions, but in the limit identical functions imply identical physical realizations. Therefore functionalism is true in approximate, but not in an absolute sense.
2. Many thought experiments in the philosophy of mind do not make sense since they are based on hidden assumption that functionalism is true in an absolute sense. In particular small changes in physical architecture do add up creating differences in functions.
3. Once such thought experiments are rejected hard problem of phenomenal experience becomes a question of finding an approximation to brain-like computing that converges to mind-like behavior.
4. Claims of qualia are necessary consequence of brain-like organization of computations, in particular the ability to comment upon physical states of the architecture carrying out these computations.
5. There is no special status of qualia or conscious experiences.

Qualia result from our ability to memorize and discriminate between real states of the global dynamics of the brain. There is nothing mysterious about them. It should be possible to construct artificial minds that will also claim qualia. Since artificial minds will not be identical to biological, because only a rough functional approximation of the brain states is possible, their qualia will be different than ours. But does it really matter? The real

questions will be of ethical nature. Something matching the complexity of human brain and adapting to its environment will have high value to its creators and teachers, regardless of its ability to really feel pleasure and pain or just to deceive itself and others that it does.

Recent discussion in philosophy of mind shows that various people find various theories of mind satisfactory. Since many of these theories – based on dual aspects of information or quantum mechanics – do not answer any empirical questions at all I do not consider them to be good theories of mind. A good theory should provide a good approximation of brain functions and connect these functions with psychological states. Cognitive science is on a proper track to construct such theory and brain/mind models. In this paper minimal requirements for an artifact that should claim qualia were given. It remains to be seen that there is something more about the experience to explain. The illusion that someone inside us is the real subject of experience is very strong, but it is possible to go beyond it. Scientific discussions on consciousness should be based on careful observations and critical evaluation of our inner experience. This is usually not the case, since almost everybody makes casual observations on his/hers state of mind – a few recent exceptions include the neurophenomenology of Varela (1996) and articles of Shanon (1998), Novak (1996) and Shear (1996).

Ancient Indian philosophy, especially Buddhist philosophy, was based on introspection and critical reflection (cf. Novak 1996). When the mind learns how to focus attention it sees that “all skandhas are empty”, as one may read in the *Heart Sutra* (Conze 1978) written more than 16 centuries ago. Five skandhas, or mutually conditioning factors, include physical body, sensations, perceptions, impulses (dispositional tendencies) and consciousness. All these are called “empty” because they do not have permanent, independent existence. “Feeling, perception, volition, even consciousness itself”, all are called empty. If we look deeply enough everything in our mind and in the material world is constantly changing (impermanent) and is mutually dependent, everything is a part of dynamical state sustained by continuous interactions. In Theravada Buddhist philosophy mind and body are on equal footing. Although mind contents and events may in reality be just shadows of neurodynamics, psychological processes admit more fruitful analysis if mind is considered to be primary. Complexity of the brain, time it takes to grow in natural environment make it the most valuable object in the Universe. We do not need the special status of conscious experience – it is just one of the ghosts left from the past.

References

- Ayer A.J, *Philosophy in the Twentieth Century*. (George Weidenfeld and Nicolson Ltd, 1982)
- Baars B.J (1988) *A Cognitive Theory of Consciousness* (Cambridge University Press, Cambridge, MA)
- Newman J, Baars B.J (1993) ‘Neural Global Workspace Model’ *Concepts in Neuroscience* 4: 255-290
- Chalmers D.J (1995) ‘Facing up to the problem of consciousness’ *J. of Consciousness Studies* 2: 200-219
- Chalmers D.J (1996) ‘Facing up to the problem of consciousness’ *J. of Consciousness Studies* 2: 200-219
- Chalmers D.J (1996a) *The Conscious Mind: In Search of a Fundamental Theory*. (Oxford University Press 1996)
- Chalmers D.J (1997) ‘Moving forward on the problem of consciousness’ *J. of Consciousness Studies* 4: 3-46
- Clarke C.J.S. (1995), ‘The non-locality of mind’. *J. of Consciousness Studies* 2: 231-240
- Cotterill R. (1996), ‘Prediction of internal feedback in conscious perception’. *J. of Consciousness Studies* 3: 245-266
- Conze, E. (1978), *Selected sayings from the Perfection of Wisdom*. (Prajna Press, Boulder)
- Dennett, D.C. (1991), *Consciousness explained*. (Little-Brown)
- Dennett D.C. (1996), ‘Facing Backwards on the Problem of Consciousness’. *J. of Consciousness Studies* 3: 4-6
- Duch W. (1994), ‘A solution to the fundamental problems of cognitive sciences’. International Philosophical Preprint Exchange (1994), paper available at <http://www.phys.uni.torun.pl/kmk/publications.html>
- Duch W. (1996), ‘From brain to mind to consciousness without hard problems’. Sympozjum Kognitywne '96: Świadomość a Percepcja. Instytut Filozofii UAM, Poznań, 6-7 Dec. 1996, available at <http://www.phys.uni.torun.pl/kmk/publications.html>
- Duch W. (1997), ‘Platonic model of mind as an approximation to neurodynamics’. In: *Brain-like computing and intelligent information systems*, ed. S-i. Amari, N. Kasabov (Springer, Singapore 1997), chap. 20, pp. 491-512
- Eccles, J. (1985), *The Brain and the unity of conscious experience*. (Cambridge University Press)
- Freeman W.J (1996), *Societies of Brains: A study in the neuroscience of love and hate*. (Lawrence Erlbaum Associates)

- Gardner M. (1996), 'Computers near the threshold'. *J. of Consciousness Studies* 3: 89-94
- Gray, J. (1995), 'The contents of consciousness: a neuropsychological conjecture'. *Brain and Behavioral Sciences* 18: 659-676.
- Gregory R.L. (1981), *Mind in Science: A History of Explanations in Psychology and Physics*. (Penguin Books)
- Hardcastle V. (1996), 'Functionalism's response to the problem of absent qualia'. *J. of Consciousness Studies* 3: 357-373
- Humpden-Turner C. (1981), *Maps of the mind*. (MacMillian)
- Jackson, F. (1982), 'Epiphenomenal qualia'. *Philosophical Quarterly* 32: 127-136
- McCune, W. (1997), 'Solution of the Robbins problem'. *Journal of Automated Reasoning* 19: 263-276
- McGinn, C. (1995), 'Consciousness and Space'. *J. of Consciousness Studies* 2: 220-230
- McClelland J.L, Rumelhart D.E. (1987), *Parallel distributed processing*. (The MIT Press, Cambridge)
- Nagel, T. (1974), 'What is it like to be a bat?' *Philosophical Review* 4: 435-50
- Nevell A. (1990), *Unified theories of cognition*. (Harvard University Press, Cambridge, Massachusetts)
- Novak, P. (1996), 'Buddhist meditation and the consciousness of time'. *J. of Consciousness Studies* 3: 267-277
- Penrose R. (1994), *Shadows of the mind* (Oxford University Press)
- Putnam, H. (1975), 'The meaning of <meaning>'. *Minnesota Studies in the Philosophy of Science* 7: 131-193
- Putnam H. (1975), *Mind, language and reality*. (Cambridge University Press)
- Putnam, H. (1987), *Representation and Reality*. (MIT Press)
- Putnam H. (1995), 'Review of *Shadows of the mind*, by Roger Penrose'. *AMS Bulletin* 32: 370-373
- Putnam H (1998) Appendix to "The Royce Lectures in the Philosophy of Mind", delivered at Brown University.
- Ripps L.J, Shoben E.J. and Smith E.E. (1973) 'Semantic distance and the verification of semantic relations'. *Journal of Verbal Learning and Verbal Behavior* 12: 1-20
- Ruppin E. (1995), 'Neural modeling of psychiatric disorders', *Network* 6: 636-656
- Sacks O. (1995), *An anthropologist on Mars* (Vintage Books, New York)
- Shanon, B. (1998), 'What is the function of consciousness?' *J. of Consciousness Studies* 5: 295-308
- Searle J. (1980), 'Minds, Brains and programs'. *Behavioral and Brain Sciences* 3: 417-458
- Shear J. (1996), 'The hard problem: closing the empirical gap'. *J. of Consciousness Studies* 3: 54-68
- Shear J. ed. (1997), *Explaining Consciousness: The Hard Problem*. (MIT Press 1997)
- Ullman S. (1996), *High level vision: Object recognition and visual cognition*. (MIT Press)
- Varela, F. (1996), 'Neurophenomenology: A methodological remedy for the hard problem'. *J. of Consciousness Studies* 3: 330-349