Daniel Kostic

Written sample

The Turing's Test and the Zombie Arguments

Belgrade
September 2004.

The Turing's Test and the Zombie Arguments

An abstract

In this paper I shall try to put some implications concerning the Turing's test and the so-called Zombie arguments into the context of philosophy of mind. My intention is not to compose a review of relevant concepts, but to discuss central problems, which originate from the Turing's test - as a paradigm of **computational** theory of mind - with the arguments, which refute **sustainability** of this thesis.

In the first section (Section I), I expose the basic computationalist presuppositions; by examining the premises of the Turing Test (TT) I argue that the TT, as a functionalist paradigm concept, underlies the computational theory of mind. I treat computationalism as a thesis that defines the human cognitive system as a physical, symbolic and semantic system, in such a manner that the description of its physical states is **isomorphic** with the description of its symbolic conditions, so that this **isomorphism** is semantically interpretable. In the second section (Section II), I discuss the Zombie arguments, and the epistemological-modal problems connected with them, which refute sustainability of computationalism. The proponents of the Zombie arguments build their attack on the computationalism on the basis of thought experiments with creatures behaviorally, functionally and physically **indistinguishable** from human beings, though these creatures do not have phenomenal experiences. According to the consequences of these thought experiments - if zombies are possible, then, the computationalism doesn't offer a satisfying explanation of consciousness. I compare my thesis from Section 1, with recent versions of Zombie arguments, which claim that computationalism fails to explain qualitative phenomenal experience. I conclude that despite the weaknesses of computationalism, which are made obvious by zombie-arguments, these arguments are not the last word when it comes to explanatory force of computationalism.

SECTION I

The imitation game

In 1950. Alan Turing asked the following question which had been of a central importance to contemporary philosophy of mind: "Can machines think?" (Turing, A.M., 1950). Discussing this question, can hardly be of any interest (and it can hardly give any fertile basis) if we don't want an answer that sounds like a techno-fair sensationalistic attraction. The above-formulated question needs to be specified. To make that question epistemologically interesting and significant we have to make it clearer and we have to define our problem precisely. Turing had supposed that he couldn't rely on usual, everyday use of the terms "machine" and "thinking". Instead of redefining the meaning of these words, he suggested some more appropriate questions, which would express and emphasize the problem in more precise fashion. Thus, instead of the question "Can machines think?" Turing formulated another, another related question: "Can machines pass a behavioral-functional test of intelligence?" The test should be an empirical test, which could offer strong evidence that machines can think. The significance of this question and its answer will be seen only if we expose the test itself and only if we make the sole question more self evident. Turing derived his test from an entertaining game named "the imitation game", in which the interrogator must guess the sex of the other two players, a male and a female. The game is set up in such a way so that the interrogator cannot make a right identification by recognizing the voice or the appearance of other participants in the game. They are located in another room so that they cannot answer the questions directly and immediately. Let us name the participants in this game in the following way: a man (A), a female (B) and the interrogator (C) – (the interrogator can be either male or female). The sex of the participants is designated as X and Y, so, at the end of the game, the interrogator has to give answers such as "A is X, and B is Y" or "B is X, and A is Y".

The answers should be written, or better, typed, so that the participants' voices would not give any clue to the interrogator. In an ideal arrangement, in Turing's time, the most appropriate was a teleprinter by which the interrogator would communicate with A and with B. However, a mediator could repeat the questions and the answers instead.

One of the participants can assist the interrogator by giving the true answers, while the other participant can perplex the interrogator by pretending that he-she is of the opposite sex. The role of

the participant B is to assist the interrogator. Her strategy is to give only true answers; she can even add different statements to her answers, like: "I am a female, don't listen to him!" However, she will not be of a significant help if a man, too, can give similar answers. (Turing, A.M. 1950.).

Turing suggested to put a machine in the place of one participant and let them, basically, play the same game. If the interrogator cannot make a difference between a machine and a human being, then we can take that as strong evidence that machines can think.

The imitation game with the machines

Turing thought that the machines involved in this game should be of a very special kind. They should be discrete state machines - the transformations of their **internal** states should be **finite** and **sudden**; they should have an infinite memory, so operations could be performed in a reasonable time, and they should be universal, so any machine with such descriptions can be involved in a test. Turing named these machines digital computers. Every one of these conditions is the key one for the thesis about the implicit **computational** assumption in the TT.

Digital computers have three basic parts:

- a storage (a memory)
- an executive unit (a processor, a processing unit, an executing of operations)
- a control unit.

Information - rules, according to which operations are performed and executed, are stored in the memory.

The executive unit performs different individual operations, which are included in computations. These individual operations may vary in individual machines. The operations may be like the following one: "Multiply 3540675445 with 7076345687", but, in some machines, they can be very simple, for example "Write 0". (Turing, A.M., 1950.).

"The Book of the rules" is in the storage. In this case, its name is "table with instructions". Its role is to control whether these instructions are performed correctly and in the right order. The control is set up in such a manner that it operates automatically. It means that it controls every operation, simultaneously with its performing.

Information in the storage are usually divided into small **packets**, which means that in a single machine a packet can contain ten decimal numbers. These numbers designate parts of the storage

in which different packets of information are put in some systematical way. A typical instruction can look like this:

"Add a number, stored in a position 6809 to the number stored in a position 4302 and put the result on a next position in the storage."(Turing, A.M., 1950.). It is needless to say that the instruction will not appear in the machine on English or any other natural language. The instruction would, more probably, be 6809430217. In this case, number "17" tells which of the different possible operations will be performed on these two numbers. In our case, the described operation is - "add number…" The control will, normally, start to perform instructions according to the sequence of positions on which instructions are stored, but, sometimes, an instruction such as "Now perform an instruction which is stored on a position 5606, and keep performing from that position" can be redirected. Or, as in this one: "If a position 4505 contains 0 perform next instruction, stored on 6707, otherwise keep performing on" (Turing, A.M. 1950.).

The instructions of this type are very important; they enable the repetition of sequences of operations on and on, until new instruction appears. We can illustrate this principle by an example from an everyday life. Tomas's mother wants Tomas to visit a shoemaker on his way to school, in order to check out whether her shoes have been repaired; to make him do so, she can either remind him of his task every day, or, she can attach a note with a reminder in a hallway, which Tomas sees before he leaves to school. When the shoes are finally repaired and Tomas brings them home, he can then destroy the message. The message in the hallway stands for the book of the rules. On the basis of the book of the rules it is possible to integrate a random element into the operating of a digital computer - it enables the digital computer to reject or destroy instructions. Even today computers are operating according to this simple principle.

Since we have described the physical structure of the machines, we can, now, begin with the exposition of the principles upon which these machines are functioning. One of the central conditions of operating is discrete states. They are immeasurably important, because, they enable one to perceive the descriptions of physical and symbolical states as identical. It also exposes the computational thesis, which underlies TT, in an obvious way.

Discrete states

We can describe digital computers as discrete states machines. These machines go from one state into another in sudden jumps. These states are different enough to avoid and ignore a possibility of confusion. An example of a system with discrete states is a light-switch. Every switch has to be either turned on or turned off. In his 1936 paper Turing (Turing, A.M., 1936.) gave a description of such machine. As an example of a discrete state machine, we can take a following system: imagine there's a ribbon on which certain symbols are printed, on a certain distance; above that ribbon there is a head, with a reader, and that head is moving above the ribbon and "reads" the symbols. The reader marks momentary position on the tape; it "reads" the symbols under, and it is one of the finite states of the machine. In every step in which the head with the reader is moving, the machine is reading the symbol, which lies on a momentary position on the ribbon. For every combination of the momentary states and the read symbols, this mechanism determines a new internal state, which can be one of these: writing symbols on the ribbon; directing of the head with the reader moves (left or right) or stopping the head. In specific case, the machine writes the symbol on the ribbon and moves one - step in the right. This procedure can be set up in such manner so the state in which the symbol is writing follows the state of moving of the head. Therefore, this machine has three discrete states and they are: 1) the state of writing the symbol on the ribbon, 2) the state of moving towards a next symbol, 3) the state of being still. A set of symbols on the ribbon consists only of "0" and "1" and the machine is set up in such a way so it begins from the outermost left symbol on the ribbon. (Turing, A.M., 1936.).

This machine can formally be described in these terms: the **internal** state of the machine (described by a position of the reader in relation to the symbol on the ribbon) can be q1, q2 or q3; an input signal can be i0 or i1 (a position in which the head reads the symbol. **Internal** state is, in every moment, determined by the last state and the input signal, as shown in this table:

Internal state	q1...q2...q3
i0(last state)	q2...q3...q1
Input (reading the symbol)	
i1(following state)	q1...q2...q3

The output signals, the only visible indicators of **internal** states (for example, the symbols written by the head on the out coming ribbon, which we can read), can be described in next table:

The state q1…q2…q3
The output 0o…0o…01 (Turing, A. M., 1950.).

This is a typical example of the discrete-state machines. They can be described with such tables; these tables show that these machines can only have **finite** number of possible states.

On the ground of the beginning state of the machine and the input signals it is always possible to predict every **following** states; on the basis of information about the input we can know every forthcoming state (output).

The description of these machines is very important, because it allow us to perceive them as physical, symbolical systems which can be described on physical, symbolical and semantic level in such a manner that their symbolical states are **instantiated** by their physical states and their symbolical states are semantically interpretable. Firstly, since it is a physical system, then there must be an adequate description of its states and their **changes** in a physical vocabulary. Thus, in principle, according to the physical laws we can predict their changes. In addition, the physical states (at least some of them) **instantiate** some constellations of symbols. The changes of these states are describable as formal or syntactic operations (computations) on the constellations of symbols. And, in the third place, the syntactic relations of symbolic states are isomorphic with semantic relations between propositions in a given language, and in that case, symbolic states are interpretable as a processing of semantic properties of the propositions of that language. So, the changes of the symbolic states in the system can be described in the terms of the semantic relations between interpreted propositions, i.e. the system can be described as if it makes conclusions, decisions, finds adequate moves in chess etc. (Buller, D.1993.). The physical description would consists of the positions of the head with the reader on the ribbon, the symbolic description would consists of the symbols on the ribbon and the logico-mathematical operations which they represent, while the semantic descriptions would consist of linguistic interpretations of output symbols. This interpretation of the discrete machines can be understood as the explicitly exposed computational thesis (Buller, D.J. 1993.). In the next section of this paper, I shall say something about other conditions and features of these machines which make them competent for the imitation game.

The Turing machine

Most digital computers have finite capacity storage. There is no theoretical difficulty in conceiving a computer with, potentially, infinite storage capacity. In the above-described example, it would be a machine, supplied with a ribbon of infinite length. Naturally, in a single operation only one single finite part of the ribbon can be used. So, only one, finitely big storage can be constructed, but we can conceive that it is possible to add more of them, if necessary. Such imaginary computers satisfy the condition of *universality*. This special feature of digital computers - so they can emulate any discrete-state machine – is extremely important, when we consider them as an universal machines. The existence of the machines with such features has a very important consequence: it is unnecessary to construct different new machines, which would perform different computing processes. All of these processes can be performed by only one single computer, adequately programmed for every single case. It can be said that it means that all digital computers are equivalent. Importance of this feature is immeasurably for the thesis of isomorphism of the description of the physical and the symbolical states, because, even a meat-chopping machine is an example of the machines with discrete states (a wheel is either turning or not turning), and it works on the input-output principle, but it is not of any interest - because it does not operate with symbols, but with meat.

This thesis enables every machine, which can be described in such a way, to enter the test. We will call these machines the computers of infinite capacity (Turing, A. M. 1936) or **Turing machines**.

On the ground of a specifically detailed description of the digital computers, which is adequate for defined features of the machines of discrete states, we can assert that the digital computer of adequate speed can imitate any discrete state machine. The imitation game can be played with these machines and the interrogator will not be able to make a difference between machines and human beings.

We can, now, consider the question from the beginning of this paper. It has been suggested to alter the question "Can machines think?" with another, more specific:” Is there a computer whose behavioral-functional performance in the imitation game interrogator could not be able to distinguish from human being?” We can make this question more general - "Are there such discrete-state machines which could fulfill these conditions?” In the context of feature of universality, we saw that both of these questions are equivalent to the next one - "Let's take a digital computer C. Is it true

that, by modifying this computer, so it has an adequate storage capacity, which increases its speed, and makes all of it possible by an adequate program, C can be constructed in such a way so that it can play a role A in the imitation game while role B would be played by a human being?"

Turing's answer to this question is affirmative. He has even exposed a few of possible objections to this thesis, saying that these objections consider only a possibility of construction of such machines and not on a significance of the test itself. The problems which come out of the Turing test are, mostly, traditional epistemological problems and they should give a sort of heuristic principle for solving of these problems. The key assumption which is the foundation of the whole test is the thesis of computationalism, as I have explained it in the section on discrete states.

The computational thesis is, actually, an explanatory model which should solve two central epistemological problems, which lye in the test itself. The first one is the problem of other minds. The problem is: the system is passing the test only if an observer thinks that he has certain mental states; but, the interrogator can acknowledge that only if he becomes the examined machine. How to avoid agnosticism and solipsism when we are faced with this problem? The second problem is a syntactic/semantic one and it opens a few more questions. Is syntax sufficient for the semantics? What is the understanding of the meaning? How can we build a system of manipulations with symbols, which would make semantics possible?

Turing test and computationalism

Central point of the Turing's test, which grounds the fields of AI and computationalism, says that we can use machines as an explanatory model for consciousness, only when the test shows that it is a successful heuristic principle. The test and the machines included, do not give any model of consciousness or intelligence on their own. Only when they pass as the heuristic principle we can use them as experimental and explanatory models and only on these ground we can build a "hard" science, i.e. the science whose theories are empirically verifiable. This is the central argument of the computationalism. It is also a central problem, as we will see in the part on the zombie argument.

Beside the problem and its consequences, with which we are faced by the TT, and which have been exposed above, TT lies on an implicit computational assumption both as the heuristic principle and as a central problem of AI (Block, N.1986., Chalmers, D. 1993. b).

The scenario of the Turing test has been set up as a behavioral-functional test. Nevertheless,

on the ground of detailed description of the test and the machines involved in it, we discover a significant implicit thesis. It is the computational thesis. Let's ask ourselves, now, in which particular aspects the exposing of the TT is, actually, the exposing of the computational thesis. In the first place, we have to mark key points in the computational theory of mind.

The basic computational thesis is the one which claims that the human cognitive system is a symbolical physical system (Buller, D.1993.). This definition consists of three levels. Firstly, as a physical system it must have an adequate description of its states and their transitions in the vocabulary of physics so, on the ground of physical laws we can predict their changes. Secondly, it is such a system in which its physical states **instances** certain constellations of symbols and the changes of these states are describable as formal or syntactic operations (computations) on these constellations of symbols. The third - in this system, when syntactic relations of their symbolical states are isomorphic with semantic relations between propositions of a certain language, then, these symbolical states are interpretable as a processing of semantic properties, as in the case with the propositions in that language. So, the changes of their symbolical states in the system can be described in the terms of semantic relations between interpreted sentences, i.e. the system can be described as if he makes conclusions, decisions, as if it makes optimal moves in chess. (Buller, D.1993.).

We can derive the epistemological importance of the computational thesis from this description. According to Chalmers, it can be expressed in two basic theses (Chalmers, D.1993.b). The first one is the thesis of *computational sufficiency*, by which it is claimed that the right kind of computational structure is sufficient for consciousness - and for capturing very different mental states. The second is the thesis concerning *computational explanation*. It claims that the computational model of consciousness gives a general frame for the explanation of cognitive processes and behaviors. Can we apply these theses on TT? We certainly can. What is estimated in the test is a specific form of behavior. We can understand this better if we take a look at the description of the test. Turing claims that machines involved in the test should have certain physical structure. The physical structure should enable a certain kind of manipulation with symbols; this manipulation is syntactic, in such a way that it is semantically interpretable. The semantic interpretation of the manipulation with the symbols and the physical states of the machine is, according to Turing, isomorphic. Isomorphism of the semantic interpretation is grounded on a fact that the machines are, actually, the systems with discrete states, so, on the bases of input (no matter if we perceive it as symbolic or as a constellation of physical states) we can predict an output. This conclusion goes far over behaviorist interpretation of TT, because it enables us to

explain un-observable internal states. The computationalism is convincing because it is a model of explanation and not a descriptive model. The consequence of such conception is that the constructing of machines which would fulfill all conditions would refute or confirm computational theory, and, it, further, means that we could set the whole field of philosophy of mind and the theories of consciousness as empirical disciplines.

In the next section, we will see to what extent is this optimism justified, especially if we expose the most problematic question in every theory of mind to a critical attack - an explanation of qualitative, phenomenal consciousness.

SECTION II

Zombies against computationalism

The proponents of the Zombie arguments claim that, if zombies are possible, then materialism, generally taken, fails. In attempting to refute materialism, especially functionalism and computationalism, they gave several thought experiments and arguments. There are many versions of this thought experiments, but we will deal only with zombie arguments as the most developed variant. Zombies have been conceived as anatomically - physiologically, and behavioral- functional identical with human beings, with the only difference that they don't have consciousness. Namely, if there are creatures that are anatomically -physiologically - behaviorally identical with us, lacking only consciousness, then, their possible existence is a counterexample to the materialistic thesis that all mental properties (consciousness) are explainable in the physics vocabulary.

In this section we will see in which manner these arguments should make this special variant of materialism, computational functionalism, weak. The efficiency of this argument is based on a fact that materialism (e.g. Physicalism) claims that there isn't anything over and above the totality of basic physical facts. If we consider this thesis in the light of computational conceptions, we will see that it can be redefined in terms of the following one: there isn't anything over and above cognitive functions, which can be described physically or as symbolic operations. This assertion allows us to interpret computationalism as a non-essentialist thesis about supervenience of mental properties: phenomenal mental states supervene on the basic physical and symbolical states. If we accept this

thesis as a true one, then the possibility of the zombies reveals an opposite example to the thesis of the supervenience of mental states. The general structure of zombie argument goes as follows:

a) If zombies are possible then the computationalism fails;
b) Zombies are possible;
c) So, the computationalism fails (Marton, P. 1988.).

Let's take a closer look into this argument.

Functionalists (computationists) claim that:
a) All facts about (phenomenal) consciousness logically supervene on the totality of basic (micro) physical facts and symbolical operations which they instantiate;
b) The proponents of zombie arguments claim that zombies, physically and behaviorally indistinguishable from human beings, although without mind, are logically possible.
c) They defend this thesis by relying on a modal principle that conceivability is sufficient for possibility;
d) so, if zombies are possible, then there are facts which do not supervene on the basic physical facts and symbolical operations which they instantiate, so, the basic functionalist (computational) thesis about supervenience is inconsistent (Marton, P, 1998.).

This conclusion corresponds with everyday thinking; if in a zombie world everything is identical with the actual one, how can our identical duplicates can be different from us only in lacking consciousness? Furthermore, it means that we have to reject our thesis of the supervenience of the facts of higher level on the basic, physical facts, because a thought experiment reveals that there are some facts of higher level which may not supervene on the basic facts.

So, if we accept the modal principle that the hypothesis is conceivable if and only if it does not contain any explicit contradiction, then we have no reasons not to accept modal conclusions in zombie arguments. But, can we, without any difficulty, accept this modal principle? Before we try to answer this question, we should remember that our thought experiment demands something more then logical condition of non-contradiction. In the scenario of the experiment there was another condition saying that zombies should not only be behavioral-functionally and physically identical with us, but, also, that they should keep all causal relations. What is that we are looking for, that exceeds the logical condition of non-contradiction? It is a causal role of consciousness, which we

want to implement into the thought experiment.

But, in all thought experiments with zombies, we suppose that phenomenal experience has a causal role. Maybe we should suppose the opposite - that phenomenal experience has no causal role. In this case our argument would turn out to be *reductio ad absurdum*. If we leave this argument without this change, there is a dangerous possibility for falling into vicious circularity. Maybe we can avoid this *circulus vitiosus* if we justify our assumption about causal role of phenomenal experience in thought experiment with the zombies. Can we do that? Philosophers who are inclined to zombie arguments would, probably, claim that we can. But, in which way?

They see a solution in a fact that this assumption needs not to be justified if we rely on a principle, which says that conceivability entails possibility. If one, on the ground of this principle, proves that zombies are possible, then the argument against computationalism is valid.

Most philosophers, who propose zombie arguments, think that conceivability is sufficient for possibility, i.e. they use notions of conceivability and possibility as if they are synonyms. Even those philosophers who make difference between special types of logical non-contradiction in this principle accept the thesis that the conceivability is sufficient for possibility (Polger, T. 2000.).

In discussions on epistemology of modal logic, opinions related to principle of conceivability, which is sufficient for possibility, are divided. As Sören Häggqvist write, many people will say that it is unconceivable that a cube will pass through a hole in the other smaller cube; or, that one can put a half spoon of a sugar in almost full glass of water without spilling the water out of the glass (Häggqvist, S, 1996, pp.127-8). Nevertheless, as we saw that sometimes, it is possible in magic tricks, though it is regarded as inconceivable.

On the other side, some things are not just possible, but even real, although they are not conceivable, i.e. they are possible with no regards to conceivability. If we consider these assertions in the light of results in neuro-science investigations on pain, we shall see how the habit to rely exclusively on conceivability is unjustified. Before we expose certain examples which are crucial for the relation between conceivability and possibility, I shall give a few introductory explanations on these explorations, so the application of these examples would be clearer.

Neuro - scientific researches on pain, in second part of the 20th century, revealed that pain, on the most general level, has two components: sensor and affective. Sensor aspect of the pain is related to neurophysiologic structures and mechanisms of our body, and these structures and mechanisms are agents of different kinds of painful sensations. The affective aspect is represented by qualitative mental states, which seems to be in a correlation with sensor structures and mechanisms (Grahek, N. 1993.). We said that these two aspects "seem" to be in the correlation,

because, it seems that in all "normal" cases physiological state of pain is followed by a certain unpleasant mental state. What makes this problem philosophically interesting is the question of a causal role of painful sensations in our phenomenal experience.

N. Grahek (Grahek, N., 2001.) gave a special interpretation of different cases of neurological disturbances, which throws a completely new light on the problem of conceivability / possibility. He emphasizes that sensory and affective components of the pain are deeply connected in “normal” circumstances. This connection is compatible with our modal intuitions. The most of us will think that painful sensations in every conceivable world should be followed with mental states, colored with unpleasantness. However, when we consider severe cases of cingulotomy and prefrontal lobotomy[1], on one hand, and cases of congenital analgesia[2]**,** on the other hand, and, then, the basics of our modal intuitions become questionable as we shall se in the following paragraph.

How can we understand reports on patients in unbearable physical pain, but, which, after a neuro-surgical operation (cingulotomy or prefrontal lobotomy) still feel the same pain, but it doesn’t worry them anymore, it doesn’t terrify them? Does that mean that phenomenal experience of the pain has been removed by the neuro-surgical intervention? On the other hand, the cases of congenital analgesia and asimboly[3] suggest a similar conclusion. We are in front of the traditional philosophical problem of causality. In these cases, we should be very careful; let’s recall different examples from the history of science in which, on the ground of physical data, it was built a completely wrong causal chain. For example, an air-pressure and “horror ex vacuo”. Grahek suggests that, when it comes to cases like these, the phenomenal experience still exists, though it is different than ours.

Among philosophers, there is a usual opinion that painful sensations are always followed by unpleasant phenomenal experiences and that it isn’t neither conceivable nor possible for them to appear without one of them. The result of pain researches reveals that it doesn’t have to be the case. The consequences of these researches should not lead us to a conclusion that the unpleasant phenomenal experiences are separable from painful sensations; but, they are, only, different. This is enough to show that even what we thought was inconceivable, is actually possible, even more – it is real, and that the principle of conceivability/possibility is not the last standpoint in the dispute between computationists and the proponents of zombie argument.

Even if we solve the problems sketched above, still there is a question about intersubjective agreement inside this principle. How to react in a case when someone claims that what we consider easily conceivable isn’t conceivable at all. As long as we stick to the first person perspective principle (and we don’t see a way to escape that principle) this dispute will remain open. Until these

disputes are solved, it seems that conceivability is not a key for constructing of thought experiments (Häggqvist, S. 1996, pp128.).

Computationalism and zombie arguments in a new light

Strictly speaking, the proponents of zombie arguments, i.e. the philosophers which believe that qualitative, phenomenal experiences are intrinsic properties of consciousness, ground their attack on computationalism on assumption which says that explanation of mind, based on the input, output and internal physical-symbolical states analysis does not give satisfying theoretical directions in the case of the explanation of consciousness. They build this assertion on the fact that the existence of zombies would not make any difference towards the computational theory of mind. On the other hand, the computationalists think that cognitive and computational functions are connected with the consciousness contingently. They think that the connection between input and behavior is more formal then causal (Moody, T.C.1994.) or they think that the causal relation can be reduced to formal. In the light of everything said in this paper, it seems that zombies are possible as long as the connection between observable states of any kind and mind is contingent (Moody, T.C., 1994.). This, actually, means that we can consider zombie-problem as an alternative formulation of other minds problem. More precisely, if zombies are possible, then, the computational criterion for attribution of consciousness stands as unsolvable problem of other minds.

The proponents of zombie arguments argue that computationalism is, actually, inessentialist theory of mind (Moody, T.C., 1994.). In light of this remark, according to computationalism, any given behavior can appear without the consciousness as a causal agent of that behavior. The only reason why would one suppose that for the explanation of certain behavior it is necessary to apply to the consciousness is that it seems like the given behavior demands a certain mental activity. If, according to computationalism, not any single explanation of mental activities demands assumption of the mind, so, it is not demanded for the explanation of behavior (Moody, T.C., 1994.). So, if the computational theory of mind is true, then zombies are possible. What are the consequences of this conception? This question puts us back to what we said in a section above. The existence of zombies would erase a difference between conscious and unconscious behavior; it would make computational thesis of sufficiency of cognitive - symbolical operations for explanation of mind insufficient [4]. This, actually, means that if computationalism is true, then we don't have any valid

criterion for difference between conscious human beings and zombies, so if zombies are possible, then, computationalism doesn't explain the consciousness.

But, if we put it in other way, the existence of zombies would, actually, be the strongest reason to accept computationalism, because, if we don't have any other criterion by which we could separate and differ conscious beings of zombies, then, the computationalism will be as close as we can ever get to explaining the consciousness.

Notes

1) Neuro-surgery operations with cutting of certain neural connections in brain (Grahek, N. 2002).
2) Native insensitiveness on pain (Grahek, N. 2002).
3) Lack of "normal" symbolic interpretation on painful sensations (Grahek, N. 2002).
4) See chapter "Turing test and computationalism".

References

1. Bringsjord, S. (1994) Precis of: What Robots Can and Can't Be. Psycoloquy 5(59).
2. Bringsjord, S. (1996) The Inverted Turing Test Is Provably Redundant. Psycoloquy 7(29).
3. Bringsjord, S. (2001) The Zombie Attack on the Computational Conception of Mind, Philosophy and Phenomenological Research, LIX.1, pp. 41–69.
4. Block, N. (1981) Psychologism and behaviourism. Philosophical Review 40, pp. 5-43.
5. Block N. (1986) The Mind as the Software of the Brain, An Invitation to Cognitive Science, edited by D. Osherson, L. Gleitman, S. Kosslyn, E. Smith and S. Sternberg, MIT Press, 1995.

6. Buller, D. J. (1993) Confirmation and the Computational Paradigm (or: Why Do You Think They Call It Artificial Intelligence?), Minds and Machines 3, pp. 155-181.
7. Chalmers, D. (1993 a). "Self-ascription without qualia: A case-study", Behavioral and Brain Sciences 16, pp. 35-36.
8. Chalmers, D. (1993 b) A Computational Foundation for the Study of Cognition, Minds and Machines (1994).
9. Chalmers, D. (1995). "Absent qualia, fading qualia, dancing qualia", Conscious Experience, ed. Thomas Metzinger. Imprint Academic, 1995.
10. Clark, A. (1987) From Folk Psychology to Naive Psychology, Cognitive Science 11, pp. 139-154.
11. Collins, H. M. (1997) The Editing Test for the Deep Problem of AI, Psycoloquy: 8 (1).
12. Cottrell, A. (1996). " Sniffing the camembert: On the conceivability of zombies, Journal of Consciousness Studies, 6, No. 1, 1999, pp. 4–12.
13. Dennett,D. (1995). "The unimagined preposterousness of zombies", Journal of Consciousness Studies, vol. 2, no. 4, 1995, pp. 322-326.
14. French, R. M. (1996) The inverted turing test: a simple (mindless) program that could pass it, Psycoloquy: 7 (39).
15. Flanagan, O. & Polger, T. (1995). "Zombies and the function of consciousness", Journal of Consciousness Studies, 2 (4), pp. 313-321.
16. Gunderson K. (1994) Movements, Actions, the Internal, and Hauser Robots, Behavior and Philosophy, 22, 1, 29-33.
17. Grahek, N. (1993) Senzorne i afektivne komponente bola, Theoria, Br. 2, str. 7-21.
18. Grahek, N. (2001) Feeling Pain and Being in Pain, BIS-Verlag, Universitaet Oldenburg.
19. Hauser, L.(1994) Acting, Intending, and Artificial Intelligence, Behavior and Philosophy, Vol. 22, No.1, pp. 22-28.
20. Hauser, L. (1993) Reaping the whirlwind: reply to Harnad's "Other bodies, other minds." Minds and Machines 3(2), pp.219-237.
21. Hauser, L. (1995). Revenge of the Zombies, American Philosophical Association Eastern Division Colloquium: Philosophy of Mind, December 29.
22. Harnad, S. (1990) The Symbol Grounding Problem. Physica, D 42, pp. 335-346.
23. Harnad, S. (1991) Other bodies, other minds. Minds and Machines 1 (1), pp. 43-54.
24. Harnad, S. (1992) The Turing Test Is Not A Trick: Turing Indistinguishability Is A Scientific Criterion, SIGART Bulletin 3(4), pp. 9 - 10.

25. Harnad, S. (1995). "Why and how we are not zombies", Journal of Consciousness Studies, 1, 164-167.
26. Harnad, S. (2001) Minds, Machines and Turing: The Indistinguishability of Indistinguishables. Journal of Logic, Language, and Information (special issue on "Alan Turing and Artificial Intelligence").
27. Häggqvist, S. (1996) Thought Experiments in Philosophy, Almqvist & Wiksell International, Stockholm.
28. Lormand, E. (1996) Nonphenomenal Consciousness, Noûs, 30, pp. 242-261.
29. Marton, P. (1998) Zombies versus Materialists: The Battle for Conceivability, Southwest Philosophy Review, 14, pp. 131-138.
30. Moor, J.H. (1976) An Analysis of the Turing test, Philosophical Studies 30, pp.249-257.
31. Moody, T. C. (1994). Conversations with Zombies, Journal of Consciousness Studies, 1 (2), pp. 196-200.
32. Nigel J.T. (1998). Zombie Killer, Stewart R. Hameroff, Alfred W. Kaszniak, & Alwyn C. Scott (eds.). Toward a Science of Consciousness II: The Second Tucson Discussions and Debates. Cambridge, MA: MIT Press, pp. 171-177.
33. Popple, A.V. (1996) The Turing Test as a Scientific Experiment. Psycoloquy 7(15).
34. Searle, J. R. (1980) Minds, brains, and programs, Behavioural and Brain Sciences 3:417-424.
35. Turing, A.M. (1950). Computing Machinery and Intelligence. In Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence, (1988). A. Collins and E. E. Smith (Eds). Kaufmann, San Mateo, CA, pp. 6-19.
36. Turing, A.M. (1936). On computable numbers with an application to the entscheindungsproblem, Proceedings of the London Mathematical Society, ser. 2. vol. 42, pp. 230-265.
37. Watt, S. (1996). Naive Psychology and the Inverted Turing Test. Psycoloquy 7(14).
38. Watt, S. (1996). A Scientific Turing test?, Psycoloquy 7(20).