Autonomy of Computation and Observer Dependence

Marcin J. Schroeder[1]

Abstract. Issue of the observer dependence of computation entered the main stage of philosophical discussion about a quarter century ago as an argument for John R. Searle's negative answer to the question "Is the brain a digital computer?" Today, this question awakens equally strong controversies in parallel with another question "What is computation?" Searle was using his own interpretation of computation, which requires involvement of a human observer, as he presented a negative answer to the first question above. This paper is a sequel to the present author's earlier paper with an answer to the second question about computation. Its main objective here is to identify what is missing in the process of computation in order to make it autonomous and suitable for modelling mind. However, autonomy is understood as independence from human intervention in the process of computation and therefore differs from Searle's "observer independence" in establishing the status of computation.

1 INTRODUCTION

Objectives of this paper may appear at first somewhat confusing, as they refer to concepts used frequently in the context of philosophy of computation and artificial intelligence (such as information, computation, etc.), but in the majority of cases they are defined in a very specific way, frequently essentially different from that of predecessors. For instance, one of my claims is that the present paradigm of computation requires an involvement of human interpreter, who may be understood as an "observer", but whose role in information integration was not considered before.

On the surface, my position may be at first mistaken for agreement with the views of John R. Searle in several of his publications, while actually my views are related to them mainly through similar terminology and in some cases are in direct contradiction to his arguments and conclusions [1-3]. For instance, Searle is using as his criterion for the objective existence of physical objects their possession of physical characteristics independent from an observer, whereas I trust that based on quantum mechanics, we know that such characteristics may be observer dependent without depraving the objects of their objectivity.

Moreover, my objectives are to identify and to present theoretical modifications of the present conception of computation which can make it autonomous and as such suitable for modelling or implementing intelligence. The first step in this direction, described as a naturalization of computation, was presented in my earlier article, where I proposed a generalization of a concept of Turing machine (symmetric Turing machine, or S-machine in distinction from Turing's A-machine) as a theoretical computing device based on processes of mutual interaction of its components (traditionally called a "head" and a "tape") instead of the orthodox one-way action of one component on the other [4]. Only in special ("artificial" or human devised) cases the interaction may be restricted to one-way actions reducing the machine to the orthodox form of an A-machine.

Motivation for such generalization was the fact that in natural processes we have always mutual interactions of components in more or less complex systems, and every one-way action is either an approximation, or is a process in which time-reversal symmetry is broken due to the hierarchical type of complexity of the system in which symmetry of lower level allows for asymmetry of the upper level (e.g. as in the case of asymmetry of the evolution of macro-states of a thermodynamic system in spite of the preserved symmetry at the level of micro-states known as the Second Law of Thermodynamics).

In my earlier paper, I was referring to arguments showing that causal explanation is our human interpretation of interaction from the first person perspective (based on our experience as agents of a very high level of complexity acting in yet more complex biological environment). Thus, every device designed according to the rules of causal relationship involve human interpretation of the process of computation. This "*Deus ex machine*" type of involvement of a human interpreter is usually obscured in the discussions of computation by the assumption that we already have a universal Turing machine and the analysis does not include the question of how this machine came to the existence.

Symmetric Turing machines are free from this assumption when put in the perspective of general dynamic of information in which they are just one of many different types of dynamic systems [5]. Actually, one of the reasons why I introduced such generalization of computation was to initiate investigation of the process in which universal Turing machines could be developed from simple dynamical systems observed in nature.

Switching from human one-way actions to natural mutual interactions of components cannot completely eliminate the presence of the human mind in computation. It remains present not in the dynamics of computation, but in the interpretation of its input and output. In my earlier description of the generalized process of computing in terms of the dynamics of information, there are two levels of its realization, local and global [5]. While at the local level, the dynamical interaction becomes entirely independent from human involvement, at the global level this involvement remains as a necessary component of computation.

[1] Akita International University, 010-1211, Akita, Japan.
Email: mjs@aiu.ac.jp.

Computation considered, for instance, as a recursive function from the set of natural numbers to itself requires an interpretation of the input or output as natural numbers, while actually on the tape there is a sequence of symbols, typically of 0's and 1's. There is nothing in the Turing machine (both the orthodox and symmetric types) which can perform the integration of information in the form of 0's and 1's into a natural number. Moreover, the interpretation is conventional. Whether 100 is a binary representation of number four or decimal representation of one hundred depends on a human interpreter.

This argument may seem very similar to those given by Searle, who is referring to the difference between the syntactic and semantic aspects of computation [1], but his arguments and his conclusions are essentially different from mine. For instance, his conclusions include the necessity of involvement of the human mind, while mine open the possibility of augmenting computation with a component performing a process of information integration freeing it from human involvement. Moreover, I am providing a theoretical description of the mechanism that can perform this interpretation [6].

Thus, there are fundamental differences between Searle's approach and my own, in terms of: conceptual frameworks, objectives, views and arguments, but his critical analysis provides a good platform for a renewed study of computation. In the following, a brief, critical summary of his views relevant to the key topics of the present paper is provided, which will prepare the stage for further discussion.

2 SEARLE'S OBSERVER DEPENDENCE

In his 1990 Presidential Address to the APA, Searle attempts to derive his conclusion that the brain is not a digital computer from the more specific thesis that computation requires an observer. He arrives at this thesis from the analysis of his definition of computation.

His first observation and at the same time first argument for his main thesis is that computation in typical attempts to define it has the feature of "multiple realizability". Computer can be made of anything. Then "[…] we get the uncomfortable result that we could make a system that does just what the brain does out of pretty much anything."[1]

This objection seems to me not very convincing. Evolutionary biology tells us that this just the way in which all functional units of leaving organisms develop from whatever is "at hand". There is no reason to believe, that human brain was an exception. Its antecedent in the early stages of evolution could have been some organizational unit of completely different function.

For Searle the "multiple realizability" is dangerous, because it implies "universal realizability". "If computation is defined in terms of the assignment of syntax then everything would be a digital computer, because any object whatever could have syntactical ascriptions made to it." [1] Then he adds "Worse yet, syntax is not intrinsic to physics." Here appears main point of his reasoning, as it is followed by "[…] the 'syntax' is not the name of a physical feature, like mass or gravity" and "[…] syntax is essentially an observer relative notion."

From this there is only one step to the thesis that computation (as that defined by the assignment of syntax) is observer relative, or in other words observer dependent. However, if the thesis of human involvement was so simple, it could be easily dismissed. The syntax of natural language is not pre-designed by someone, but is as natural characteristic of reality as mass or gravity, but at the higher level of complexity within living objects. If the involvement of human communicators in the use of language obscures the natural character of syntax in the historical development of natural languages, we can think about its early forms in the languages of bees, ants, birds or dolphins.

It seems that in the following parts of the Presidential Address as well as in other works of Searle that the actual problem is not in the opposition of computation to the objects of physics, but that the actual problem is with the mutual relationship between syntax and semantics.

Thus, the actual argument for the necessity for a human observer of computation is very similar to one of classical philosophical problems of meaning. It is not very different from Brentano's thesis that intentionality associated with the semantic relations is a main characteristic of mind or mental reality, which makes it essentially different from physical reality [7].

Searle believes that it was the Chinese Room Argument which was addressing the issue of the relationship between syntax and semantics ("semantics is not intrinsic to syntax" [1]) and that the present issue requires another thesis that "syntax is not intrinsic to physics" [1]. He writes "There is no way you could discover that something is intrinsically a digital computer because the characterization of it as a digital computer is always relative to an observer who assigns a syntactical interpretation to the purely physical features of the system […] It requires the assignment of a computational interpretation by some agent." [1]

Searle does not write what he means by "physics", but whatever was his understanding of this word, his arguments are fallacious, because physics, in any understanding of this word cannot be considered a monolith free from the same type of problems as that in the study of computation. "Syntax is not part of physics" [1] is true, but is not relevant for the discussion of the necessity of human involvement in giving a physical process the status of computation. Natural selection is not part of physics either, but we do not need human involvement in its functioning. Physical theories have their syntax, and physical empirical methods describe their semantics. Moreover, formalisms of physical theories such as quantum mechanics can be understood as alternative syntax for empirical logic, which motivated the development of so called quantum logics.

Thus, Searle's arguments do not go beyond the old problem of the meaning of meaning. The problem in the context of computation is in the involvement of human interpreter in assigning the meaning to the input and output of computation. It is not necessary to have an observer watching the work of Turing machine, as at this stage interpretation is made at the local level of its components, but the involvement of an observer, or rather interpreter becomes necessary when we want to analyse the relationship between the input and output. Here we have to be able to understand the process of assigning the meaning.

This issue can be addressed in many different ways. In my earlier paper, I proposed a solution in terms of information and its integration, which is lying in the foundations of my view on autonomous computation in the present paper [8].

The solution comes through the assumption that whenever we are using a language (natural or artificial) to express some statement about reality, we are actually making statement about information which constitutes this reality. Thus, semantics is a relationship between two information systems, that of language and that of the subject of its statements. This relation has to preserve information structures of the two systems. It could be said in short that semantics is a relationship between two syntax systems, but considered not at the level of language, but of information.

The need for semantics (or for the study of meaning) comes from the limitations of the capability to process information by human brain. Typically, a symbol has much smaller volume of information than its designate. Using language in its symbolic form, we can manipulate relatively simple symbols instead of sometimes very complex objects which they represent. The word "cow" is related to a very complex information system which is walking in the fields. Multiple realizability of the cow is not a problem, as long as the process of information integration into a whole is clear. Of course, this approach is very distant from the position maintained by Searle.

One point where I fully agree with Searle is that the present concept of computation by the orthodox Turing machines is haunted by the homunculus fallacy [1]. I agree with his criticism of Dennett's exorcism of the homunculus fallacy by considering a sequence of progressively stupider homunculi, but my own arguments, originally in the context of the explanation of consciousness, were different [7]. As it was mentioned above, the main issue here is that frequently it is overlooked that the input and output of computation are interpreted, for instance as numbers, but in reality they have the form of a sequence of component symbols. Without homunculus, or human observer, there is nothing which can perform integration of such distributed information into a whole.

Thus, the issue whether human involvement in computation is necessary or not can be reformulated as a question whether it is possible to develop an artificial, but autonomous system integrating information.

The following will be a short overview of the conceptual framework developed in my earlier articles necessary for the further discussion of the autonomy of computation in terms of information integration.

3 CONCEPTS OF INFORMATION, ITS INTEGRATION, AND COMPUTATION

My approach to the study of computation is based on the concept of information. At this point my views diverge completely from those of Searle. In his understanding "'Information' does not name a real physical feature of real world in the way that neuron firings, and for that matter consciousness, are real physical features of the world. Except for the information that is already in the mind of some conscious agent, information is relative to an observer […] Information is anything that we can count or use as information." [3]

I do not know what Searle means by "information that is already in the mind," but for me information does not require existence of any conscious agent, and this view is fully consistent with the dominating position in the studies of this concept. For instance, biological evolution is a process based on the dynamics of information which led to the development of consciousness. What Searle calls "information" seems to be closer to a common sense concept of knowledge, but more important is just to emphasize the difference in the use of this term in this paper.

I understand the concept of information in the way it was defined in my earlier papers [9] as an identification of a variety. Thus, starting point in the conceptualization of information is in the categorical opposition of the one and many, which is the most fundamental characteristic of physical reality. Without this distinction and its contrasting character no description of reality is possible.

The variety in this definition, corresponding to the "many" side of the opposition is a carrier of information. Its identification is understood as anything which makes it one or a whole, i.e. which moves it into the other side of the opposition. The word "identification" (instead of the simpler word "unity") indicates that information gives an identity to a variety. However, this identity is considered an expression of unity, "oneness".

There are two basic forms of identification. One consists in selection of one out of many in the variety, the other of a structure binding many into one. This brings two manifestations of information, the selective and the structural. The two possibilities are not dividing information into two types, as the occurrence of one is always accompanied by the other, but not on the same variety, i.e. not on the same information carrier.

For instance, information used in order to open a lock with the corresponding key can be viewed in two alternative ways. We can think about a proper selection of the key, out of some variety of keys, or we can think about the spatial structure of the key which fits the structure of the lock. In the first case, the variety consists of the keys, in the second the variety consists of material units forming appropriate shape of the key. Thus, we can consider selective and structural information as dual, but co-existing manifestations of one concept.

The identification of a variety may differ in the degree. For the selective manifestation this degree can be quantitatively described using an appropriate probability distribution and can be measured using, for instance, Shannon's entropy, or other more appropriate measure.

For the structural manifestation the degree can be characterized in terms of the level of decomposability of the structure [6]. The concepts of information and its integration can be formalized in a mathematical theory, although for the present paper this formalization is not necessary. It is enough to mention that decomposability of the information structures becomes familiar mathematical concept of reducibility to a direct product of any type of mathematical structures [10]. Thus, the level of information integration refers to the degree in which structures carrying information maintain their integrity.

The process of transformation of information into different levels of integration was theoretically described as a generalized Venn gate whose output can have arbitrary level of integration [6]. Its physical realization is an open question, but quantum mechanical systems give an example of complete integration.

Computation is understood as dynamical interaction of two information systems, which traditionally were called a "head" and a "tape". Each of them has two levels, local and global. Tape consists of cells, where at the local level each cell is an information system which can have one of many characters out of some alphabet (selective manifestation of information). In the head instead of cells we have instruction list positions (*ilp's*). Each such *ilp* can hold one of many instructions from the catalogue of instructions (selective information at the local level) [4,5].

At the global level we have structural manifestation of information I the form of the configuration of characters in all cells of the tape and the configuration of instructions on the list in the head.

Thus far the description is essentially the same for an A-machine and for an S-machine. Difference begins now when we assume that both cells and ilp's include in their description of the local state the information about the change in passing to the next step of computation. States of the cells are characterized by their current character and they have instruction how to change when the cell comes into interaction with *ilp* in each of possible states. State of the *ilp* is characterized by the description how it should change when it comes into local interaction with cells in all possible states.

The dynamics of interaction at the global level dictates which pairs of cells and *ilp's* will be coming into contact and interaction based on the current pair in contact. The only change in comparison with A-machines is in the possibility of changes of instructions in the head. The fact that we dissociate the process of selection of next pair cell-*ilp* from local instructions in the head is purely formal. If we assume that the local dynamics does not change instructions in *ilp's*, we get the orthodox A-machine.

4 ANALOGUE AND DIGITAL INFORMATION

The problem addressed by Searle and which is the main subject of the present paper is obscured by the common confusion regarding the division of information (or as it is usually said data) into analogue and digital. In my earlier work, I was showing that this division is not related to discrete or continuous characteristics of numbers, but to the interpretation of information [5], and therefore to the issue of the meaning of information [8].

We talk about analogue information when information is interpreted in terms of physical observables, while in the case of digital information interpretation is made in terms of configuration of digits. Actually, we can have combinations of interpretations of both types, as in an example given by Ned Block quoted by Searle "[…]we can have electrical gates where 1's and 0's are assigned to voltage levels of 4 volts and 7 volts respectively" [1]. The first step of interpretation is in terms of the empirical procedure of an identification of the values of the physical observable of electric potential, and the second step is an assignment of the values of 1 or 0. However, the values 4 and 7 are equally conventional as 1 and 0. So, there is no essential difference between the two interpretations.

Actually, the issue appears when we have, as in Turing machine, complex systems of symbols. If Turing machine had only one cell with possible characters 1 and 0, problem does not appear at all, but we do not have computation in this case. I cannot agree with Searle that window in front of him can be interpreted as a "simple computer" with open window corresponding to 1 and closed window to 0 [3].

The essence of computing is in the distinction between two levels, that of digits, and that of numerals. It is not a matter of natural or artificial intelligence. I doubt that those working in offices as (human) computers before the Second World War read the sequences of digits in the results of their work integrated into numbers. They were working at the level of digits. The same applied to the work of typists, for whom every attempt to understand the text was slowing down the work. But the process of computing was for the purpose of relating the input and output, where interpretation becomes necessary, for instance, when we think about numbers. The sequences of digits produced by the human computers were usually read by someone else, for whom these sequences had some specific meaning.

Thus fundamental question is whether this interpretation or assignment of meaning can be done by a non-human agent and whether it can be integrated into the work of computing machine. In some cases and in some ways it may seem possible in an obvious way by the incorporation of peripheral devices into computers. For instance, we can display photographs or play music on a computer. However, the actual integration of the pattern of pixels into a picture or the sequence of sounds into a melody is being made by a human agent. Computer is not viewing the picture, it is only displaying it.

We can conceive situations when peripheral devices are more sophisticated. For instance, the computer can associate the pattern of pixels in the picture of a fingerprint with the name of a person. In this case, computer is relating the structural information in the picture with the structural information in the combination of letters forming a name. The association however is done not through integration of information performed by a computer, but at the digital level, using associations prepared earlier by a human programmer.

5 AUTONOMOUS COMPUTATION AND AUTONOMOUS INTELLIGENCE

The issue of the dependence of computation on a human observer or agent is therefore basically the same as the issue of the existence of an information processing system which can integrate information, but which is not dependent on the involvement of the human agent, or rather human mind. This is the reason why I prefer the expression of "autonomous information integration" instead of "observation". The latter is not enough specific (what is "observation"?) and suggests a passive role (integration of information is not an observation, but active participation in the process).

Computation becomes autonomous, when it does not involve human information integration, of course with additional assumptions of the lack of human intervention in all other parts of the process, which can be easily achieved. We could see that for the purpose of designing such process, we have to be able to create a system which operates at the global level of computation which can interpret (i.e. can give meaning to) the structural manifestation of information in the configurations of the input and output of the processes at the local level.

In the consequence, we have to ask about feasibility to realize any system which can integrate information, different from human brain. Of course, the fact that human brain can do it is reassuring that such system can exist, but only to those who believe that mind is derived in some way from the workings of the brain.

Before we can talk about the potential existence of systems integrating information, we have to ask about the existence of instances of integrated information. The obvious example can be found in quantum physics. System in a superposition of quantum states is a canonical example of completely integrated information. In fact my idea of information integration came out of this example [6]. However, we can associate complete information integration with classical geometric systems [11]. This fact guided me in my attempts to design geometric computing machines [12].

The key point in looking for some form of hypercomputing to resolve the problem of designing autonomous computing is an old, very fundamental, but frequently overlooked result of Alfred Tarski regarding indefinability of the truth [13]. He showed that in order to consider semantics for a syntax system, it is necessary to use more powerful logical system. In translation into the present context, every computation system which can include semantics for Turing machines has to have more computational power than Turing machine.

5 CONCLUSION

Although the arguments used by Searle in presenting his view of the dependence of computation on a human observer are for me not conclusive, his objections reflect different, but equally important involvement of a human agent in the orthodox form of computation performed by Turing machines. Searle distanced himself in this context from the problem of assigning meaning, but his view is based on the fallacious concept of physics standing outside of the same type of problems as those discussed here.

Development of fully autonomous computing systems requires incorporation into their design components capable of information integration. Since Turing machines lack such capability, and no system equivalent to Turing machine can equip Turing machine with the ability to generate meaning for its language, it is necessary to look for computation in more general and more powerful form.

The description of such form of computation can be done within the framework of information, its integration, and dynamics. The next step is an implementation of a device capable to perform autonomous computation. At this point, it is too early to claim that such a device can model human brain equipped with consciousness, but this theoretical description opens the possibility that the brain is actually a natural computing system.

Acknowledgment. I would like to express my appreciation for the language corrections suggested by my colleague Don Nilson. However, I am fully responsible for all remaining errors, as well as for all opinions and theses of the paper.

REFERENCES

[1] J.R. Searle. Is the Brain a Digital Computer? *Presidential Address to the American Philosophical Association, 1990.* http://users.ecs.soton.ac.uk/harnad/Papers/Py104/searle.comp.html/ retrived 12/21/2013.

[2] J.R. Searle. *The Rediscovery of the Mind.* MIT Press, Cambridge, MA (1992).

[3] J.R. Searle. *The Mystery of Consciousness.* NYREV, New York (1997).

[4] M.J. Schroeder. Schroeder. From Proactive to Interactive Theory of Computation. In: M. Bishop and Y. J. Erden (Eds.): *The 6th AISB Symposium on Computing and Philosophy: The Scandal of Computation – What is Computation?* The Society for the Study of Artificial Intelligence and the Simulation of Behaviour, pp. 47-51, http://www.aisb.org.uk/ (2013).

[5] M.J. Schroeder. Dualism of Selective and Structural Manifestations of Information in Modelling of Information Dynamics, In: G. Dodig-Crnkovic, R. Giovagnoli (Eds.) *Computing Nature, SAPERE 7,* Springer, Berlin, Germany, pp. 125-137, (2013).

[6] M.J. Schroeder. Quantum Coherence without Quantum Mechanics in Modeling the Unity of Consciousness, In: P. Bruza, et al. (Eds*.) QI 2009*, LNAI 5494, Springer, Berlin, Germany, pp. 97-112, (2009).

[7] M.J. Schroeder. Concept of Information as a Bridge between Mind and Brain. *Information*, 2 (3), 478-509, (2011).

[8] M.J. Schroeder. Semantics of Information: Meaning and Truth as Relationships between Information Carriers, In: C. Ess & R. Hagengruber (Eds.) *The Computational Turn: Past, Presents, Futures? Proc. IACAP 2011, Aarhus University – July 4-6, 2011*. Monsenstein und Vannerdat Wiss., Munster, Germany, pp. 120-123, (2011).

[9] M.J. Schroeder. Philosophical Foundations for the Concept of Information: Selective and Structural Information. In: *Proc. of the Third International Conference on the Foundations of Information Science, Paris 2005*, http://www.mdpi.org/fis2005 , (2005).

[10] M.J. Schroeder. From Philosophy to Theory of Information*, Intl. J. Information Theor. and Appl.*, 18 (1), 56-68, (2011).

[11] M.J. Schroeder. On Classification of Closure Spaces: Search for the Methods and Criteria. In: (Ed.) Akihiro Yamamura, *RIMS Kokyuroku 1809, Algebraic Systems and Theoretical Computer Science*, Research Institute for Mathematical Sciences, Kyoto University, Kyoto, pp. 145-154, (2012).

[12] M.J. Schroeder. Computing as Dynamics of Information: Classification of Geometric Dynamical Information Systems Based on Properties of Closure Spaces. In: (Ed.) Akihiro Yamamura, *RIMS Kokyuroku, Algebraic Systems and Theoretical Computer Science*, Research Institute for Mathematical Sciences, Kyoto University, Kyoto, (to appear, 2014).

[13] A. Tarski, tr. J.H. Woodger. The Concept of Truth in Formalized Languages. (English transl. of the original 1936 article in German) in J. Corcoran (ed.) *Logic, Semantics, Metamathematics.* Hackett (1983).