

*** Published in APA Newsletter Philosophy and computers Fall 2013 ***

<http://c.vmdn.com/sites/www.apaonline.org/resource/collection/EADE8D52-8D02-4136-9A2A-729368501E43/ComputersV13n1.pdf>

Turing Test, Chinese Room Argument, Symbol Grounding Problem. Meanings in Artificial Agents

Christophe Menant

<http://crmenant.free.fr/Home-Page/index.HTM>

Abstract

The Turing Test (TT), the Chinese Room Argument (CRA), and the Symbol Grounding Problem (SGP) are about the question “can machines think?” We propose to look at these approaches to Artificial Intelligence (AI) by showing that they all address the possibility for Artificial Agents (AAs) to generate meaningful information (meanings) as we humans do. The initial question about thinking machines is then reformulated into “can AAs generate meanings like humans do?”

We correspondingly present the TT, the CRA and the SGP as being about generation of human-like meanings. We model and address such possibility by using the Meaning Generator System (MGS) where a system submitted to an internal constraint generates a meaning in order to satisfy the constraint. The system approach of the MGS allows comparing meaning generations in animals, humans and AAs. The comparison shows that in order to have AAs capable of generating human-like meanings, we need the AAs to carry human constraints. And transferring human constraints to AAs raises concerns coming from the unknown natures of life and human mind which are at the root of human constraints.

Implications for the TT, the CRA and the SGP are highlighted. It is shown that designing AAs capable of thinking like humans needs an understanding about the natures of life and human mind that we do not have today. Following an evolutionary approach, we propose as a first entry point an investigation about the possibility for extending a “stay alive” constraint into AAs.

Ethical concerns are raised from the relations between human constraints and human values. Continuations are proposed.

(This paper is an extended version of the proceedings of an AISB/IACAP 2012 presentation (<http://www.mrtc.mdh.se/~gdc/work/AISB-IACAP-2012/NaturalComputingProceedings-2012-06-22.pdf>).

1. Turing Test, Chinese Room Argument and Meaning Generation

The question “can machines think?” has been addressed in 1950 by Alan Turing and formalized by a test, the Turing Test (TT), where a computer is to answer questions asked by humans. If the answers coming from the computer are not distinguishable from the ones made by humans, the computer passes the TT [Turing, 1950]. So the TT addresses the capability for a computer to understand questions formulated in human language and answer these questions as well as humans would do. Regarding human language, we consider that understanding a question is to access the meaning of the question. And answering a question obviously goes with generating the meaning of the answer. So we consider that the TT is about meaning generation.

The validity of the TT has been challenged in 1980 by John Searle with a thought experiment, the Chinese Room Argument (CRA), aimed at showing that a computer can pass the TT without understanding symbols [Searle, 1980]. A person not speaking Chinese and exchanging Chinese

symbols with people speaking Chinese can make them believe she speaks Chinese if she chooses the symbols by following precise rules written by Chinese speaking persons. The person not speaking Chinese passes the TT. A computer following the same precise rules would also pass the TT. In both cases the meaning of the Chinese symbols is not understood. The CRA argues that the TT is not valid for testing machine thinking capability as it can be passed without associating any meaning to the exchanged information. Here also, the understanding of the symbols goes with generating the meanings related to the symbols. So we can consider that the TT and the CRA are about the possibility for AAs to generate human-like meanings. This brings the question about machines capable to think into a question on meaning generation. Can AAs generate human-like meanings?

In order to compare the meanings generated by humans and by AAs, we use the Meaning Generator System (MGS). The MGS models a system submitted to an internal constraint that generates a meaning when it receives information that has a connection with the constraint. The generated meaning is precisely the connection existing between the received information and the constraint, and it is used to determine an action that will be implemented in order to satisfy the constraint [Menant, 2003].

The MGS is simple. It can model meaning generation in elementary life. A paramecium moving away from acid water can be modeled as a system submitted to a 'stay alive' constraint that senses acid water and generates a meaning 'presence of acid not compatible with the 'stay alive' constraint'. That meaning is used to trigger an action from the paramecium: get away from acid water. It is clear that the paramecium does not possess an information processing system that would allow her to have access to an inner language. But a paramecium has usage of sensors that can participate to a measurement of the acidity of the environment. The information made available with the help of these sensors will be part of the process that will generate the move of the paramecium in the direction of less acid water. So we can say that the paramecium has overall created a meaning related to the hostility of her environment in connection with the satisfaction of her vital constraint. Fig 1 represents the MGS with this example.

The MGS is a simple tool modeling a system submitted to an internal constraint [1]. It can be used as a building block for higher level systems (agents) like animals, humans or AAs, assuming we identify clearly enough the constraints corresponding to each case [2].

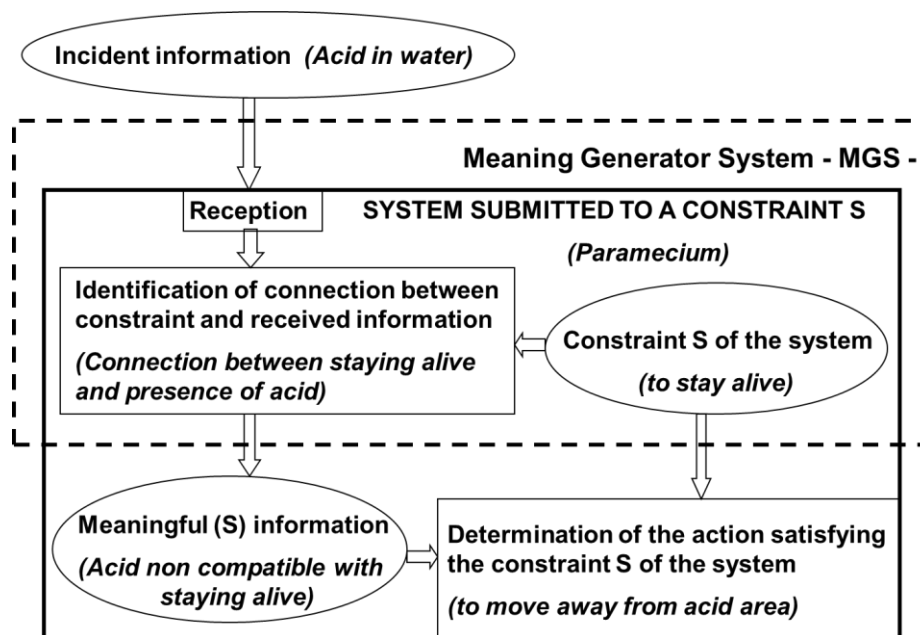


Figure 1. The Meaning Generator System

The function of the meaningful information is to participate to the determination of an action that will be implemented in order to satisfy the constraint of the system. This makes clear that a meaning does not exist by itself. A meaning is meaningful information about an entity of the environment which is generated by and for a system submitted to an internal constraint that characterizes the system.

The MGS approach is close to a simplified version of the triadic Peircean theory of sign (Sign, Object, Interpretant). Peirce's theory is a general theory of sign, and the MGS approach is centered on meaning. The MGS can be compared to a simplified version of the Peircean Interpreter producing the Interpretant. The generated meaning combines an objective entity of the environment (the incident information) and a specific construction of the system (the connection with the constraint). The MGS displays a simple complementarity between objectivism and constructivism.

The MGS is also usable to position meaning generation in an evolutionary approach. The starting point is basic life with a 'stay alive' constraint (for individuals and for species) and a 'group life' constraint. The sight of a cat generates a meaning within a mouse, as well as a passing by fly within a hungry frog. But the 'stay alive' constraint refers to life, the nature of which is unknown as of today. What can be accessed and analyzed are the actions that will be implemented to satisfy the 'stay alive' constraint, not the constraint. For humans, the constraints are more difficult to identify. They are linked to human consciousness and free will which are both mysterious concepts for today science and philosophy. Some aspects of human constraints are however easy to guess, like 'look for happiness' or 'limit anxiety' [3]. References to the Maslow pyramid can also be used as an approach to human constraints [Menant, 2011]. But what can be understood about these constraints refers mostly to the actions implemented to satisfy them. The nature of the constraints is unknown as related to the still mysterious human mind.

In all cases the action implemented to satisfy the constraint will modify the environment, and so the generated meaning. As said, meanings do not exist by themselves. They are agent related and come from meaning generation processes that link the agents to their environments in a dynamic mode. Different systems can generate different meanings when receiving the same information. And incident information can be meaningful or meaningless [4].

Most of the time agents contain several MGSs related to different sensorimotor systems and different constraints to be satisfied. An item of the environment generates different interdependent meanings that build up networks of meanings representing the item to the agent. These meaningful representations embed the agent in its environment through constraints satisfaction processes.

To see if AAs can generate meanings like humans do we have to look at how human meaning generation processes could be transferred to AAs. Fig 1 shows that the constraint is the key element to be considered in the MGS. The other elements deal with data processing that is transferrable. But when looking at transferring human constraints to AAs, we face the problem of the unknown natures of life and human mind from which these constraints result. Take for instance the basic 'stay alive' constraint that we share with animals. We know the actions that are to be implemented in order to satisfy that constraint, like keep healthy and avoid dangers. But we do not really know what life is. We understand that life came out of matter during evolution, but we do not know how life could be today built up from inanimate matter. The nature of life is a mystery. Consequently, we cannot transfer a 'stay alive' constraint to AAs because we cannot transfer something we do not understand. The same applies for human specific constraints which are closely linked to human mind. We do not know exactly what is 'look for happiness'. We only know (more or less) the physical or mental actions that should be implemented in order to satisfy this complex constraint. So we have to face the fact that the transfer of human constraints to AAs is not today possible as we cannot transfer things we do not know.

The proposed approach shows that we cannot today build AAs able to generate human-like meanings. In the TT, the computer is not in a position to generate meanings like humans do. The computer cannot understand the questions nor the answers as humans do. It cannot pass the TT. Consequently, the CRA is right. Today AAs cannot think like humans think. Strong AI is not possible today. A better understanding about the natures of life and human mind is necessary for a progress toward the design of AAs capable of thinking like humans think. Research activities are in process in these areas [Philpapers, Nature of Consciousness, Nature of life]. Some possible short cuts may be investigated, at least for the transfer of animal constraints (see hereunder).

2. Symbol Grounding Problem and Meaning Generation

The possibility for computers to attribute meanings to words or symbols has been formalized by Stevan Harnad in 1990 through the Symbol Grounding Problem (SGP) [Harnad, 1990]. The SGP is generally understood as being about how an AA computing with meaningless symbols can generate meanings which are intrinsic to the AA. "How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols?"

The SGP being about the possibility for AAs to attribute intrinsic meanings to words or symbols, we can use the MGS as a tool for an analysis of the intrinsic aspect of the generated meaning. The MGS defines a meaning for a system submitted to an internal constraint as being the connection existing between the constraint and the information received from the environment. The intrinsic aspect of the generated meaning results from the intrinsicness of the constraint. In order to generate an intrinsic meaning, an agent has to be submitted to an intrinsic constraint. Putting aside metaphysical perspectives, we can say that the performance of meaningful information generation appeared on earth with the first living entities. Life is submitted to an intrinsic and local 'stay alive' constraint that exists only where life is and is not present in the material world surrounding the living entity. As today AAs are made with material elements, they cannot generate intrinsic meanings because they do not contain intrinsic constraints. So the semantic interpretation of meaningless symbols cannot be intrinsic to AAs. The SGP cannot have a solution in the world of today AAs [5].

The same conclusion can be reached by recalling the impossibility to transfer human constraints to AAs. The constraints that are present in the AAs are derived constraints implemented by the designers (like win chess or avoid obstacles). These constraints come from the designer of the AA. They are not intrinsic to the agent like are 'stay alive' or 'look for happiness' constraints. AAs can only generate derived meanings coming from their derived constraints. Today AAs cannot carry intrinsic constraints, and consequently cannot generate intrinsic meanings. Again, the SGP cannot have a solution in the world of today AAs.

The conclusions reached in the previous paragraph apply. AAs cannot today generate meanings nor think like we humans do. We need better understandings about the natures of life and human mind in order to address the possibility for human-like meaning generation and thinking in AAs.

Another area of investigation for intrinsic constraints in AAs is to look for AAs capable of creating their own constraints. Whatever the possible paths in this area, it should be highlighted that such approach would not be enough to allow the design of AAs able to think like humans do. The constraints that the AAs might be able to generate by themselves may be different from human ones or managed differently by the AAs. These future AAs may think, but not like humans think. This brings up ethical concerns for AI where AAs would not be managing constraints and meanings the same way humans do.

3. Artificial Intelligence, Artificial Life and Meaning Generation

The above usage of the MGS with the TT, the CRA and the SGP has shown that machines cannot today think like humans do because human constraints are not transferrable to AAs. The basic ‘stay alive’ constraint is also part of human constraints, and not being able to transfer it to AAs implies that we cannot design AAs managing meanings like living entities do. Strong artificial life (AL) is not possible. So not only can’t we design AAs able to think like humans think, we can’t even design AAs able to live like animals live. At this level of analysis, the blocking points in AI and in AL come more from our lack of understanding about the natures of life and human mind than from a lack of computer performances. Progresses in AL and in AI need more investigations about the nature of life and the nature of human mind.

In terms of increasing complexity, these subjects can be positioned following an evolutionary approach. As life came up on earth before human mind, it should be easier and logical to address first the problem about the ‘stay alive’ constraint not transferrable to AAs. Even if we do not know the nature of life, we are able to manipulate it. And we could, instead of trying to transfer the performances of life to AAs, look at how it could be possible to extend life to AAs, without needing an understanding about the nature of life. In a way to be defined, we would bring the AA at the level of a living entity. We would design an agent being at the same time alive and artificial. An agent being alive (submitted to a ‘stay alive’ constraint), and being artificial (on which we keep some control). Research activities are in process on close domains like integrating the computational capabilities of neurons in robots control circuits [Warwick and al, 2010] or designing insect-machine hybrids with motor control of insects [Bozkurt and al, 2009]. These research activities are promising for the development of biological computing and life-AA merging, but the possibility for extending a ‘stay alive’ constraint to the AA is to be investigated. Such possible progress about having AAs submitted to resident animal constraints does not solution the problem about AAs submitted to human constraints, but we can however take this as a first step in an evolutionary approach to AAs containing human constraints.

4. Meaning Generation, Constraints, Values and Ethical Concerns

The MGS approach has shown that our current lack of understanding about the natures of life and human mind make impossible today the design of AAs able to think like humans do. The reason being that we do not know how to transfer human constraints (like ‘look for happiness’) to AAs. But human constraints do not a priori include human values (some humans find happiness by the suffering of others). So looking at transferring human constraint to AAs brings up ethical concerns. AAs submitted to human constraints may not carry human values. Researches about the nature of human mind and artificial intelligence should consider how human values could be linked to human constraints. It is a challenging subject because human values are not universal and human constraints remain ill defined. But the nature of human mind is still to be discovered and we can hope that its understanding will shed some light on the diversity of human values. Also, as addressed above, another case is the one about AAs becoming capable of generating by themselves their own constraints. Such approach should keep human values in the background of these constraints so the AAs are not brought to generate meanings and actions too distant from human values.

5. Conclusions

We have proposed that the TT, the CRA & the SGP can be understood as being about the possibility for AAs to generate human like meanings. Using that analogy, it has been argued that AAs cannot think like humans think because they cannot generate human-like meanings. This has been shown by using a model of meaning generation for internal constraint satisfaction (the MGS). The model shows that our lack of understanding about the natures of life and human mind makes impossible the transfer

of human constraints to AAs. Consequently, today AAs cannot think like we humans think. They cannot pass the TT. The CRA is correct and the SGP cannot have a solution. Strong AI is not possible today. Only weak AI is possible. Imitation performances can be almost perfect and make us believe that AAs generate human-like meanings, but there is no such meaning generation as AAs do not carry human constraints. AAs do not think like we do. Another consequence is that it is not possible today to design living machines. AAs cannot generate meanings like animals do because we do not know the nature of life and cannot transfer animal constraints to AAs. Strong AL is not possible today. At this level of analysis the blocking points for strong AI and strong AL come more from our lack of understanding about life and human mind than from computers performances. We need progresses in these understandings to design AAs capable of behaving like animals and thinking like humans. As life is less complex and easier to understand than consciousness, the transfer of a 'stay alive' constraint to AAs should be addressed first. An option could be to extend life with its 'stay alive' constraint within AAs. The AA would then be submitted to the constraints brought in with the living entity.

Ethical concerns have been raised through the possible relations between human constraints and human values. If AAs can someday be submitted to human constraints, they may not carry human values.

6. Continuations

The MGS approach applied to the TT, the CRA and the SGP has shown that the constraints to be satisfied are at the core of the meaning generation process and that it is not possible today to transfer animal or human constraints to AAs because of our lack of understanding about life and human mind. As a consequence it is not possible today to design AAs that can live like animals or think like humans. This status brings to consider further developments linking constraints, life and human mind in an evolutionary background.

An evolutionary approach to the nature of constraints should open the way to an understanding of a continuity of constraints from animal to humans. It would support an evolutionary theory of meaning and may provide new perspectives for an understanding about the nature of life and the nature of human mind. It may also support the possibility to address human constraints without using animal ones (i.e. address strong AI without usage of strong AL)

Identifying the origin of biological constraints relatively to physico-chemical laws may allow to start an evolutionary theory of meaning in the material world. Work is in process on these subjects [Riofrio, 2007].

The MGS approach also offers the possibility to define meaningful representations that embed agents in their environments. Such representations can be used as tools in an evolutionary approach to self-consciousness where the human constraints play a key role. Work is in process in this area [Menant, 2010].

An evolutionary approach to human constraints brings to address the 'stay alive' constraint that we share with animals. But the nature of life is today a mystery. As introduced above, we feel it could be interesting to investigate the possibility to have a living entity extending its 'stay alive' constraint within AAs. We could then have AAs submitted to the 'stay alive' constraint without needing an understanding about the nature of life.

Regarding ethical concerns, an evolutionary approach to human consciousness could introduce a common evolutionary background for constraints and values. Such concern applies also to the possibility of AAs creating their own constraints that may be different from human ones and consequently not linked to human values.

End notes

- [1] In the MGS approach the constraint is proper to the system that generates the meaning (see Fig. 1). The constraint is related to the nature of the system.
- [2] The MGS approach is based on meaning generation for constraint satisfaction. It is different from ‘action oriented meaning’. With the MGS, the constraint to be satisfied is the cause of the generated meaning which determines the action that will be implemented to satisfy the constraint. The meaning is then ‘constraint satisfaction oriented’. The action comes after [Menant, 2011].
- [3] ‘Anxiety limitation’ has been proposed as a constraint feeding an evolutionary engine that could have lead pre-human primates to the performance of self-consciousness [Menant, 2005 a,b, 2010].
- [4] Such usage of meaningful information is different from the Standard Definition of semantic Information (SDI) linked to linguistics where information is meaningful data [Floridi, 2003]. Our system approach addresses all types of meaning generation by a system submitted to an internal constraint. It covers the cases of non linguistic meanings (animals and AAs).
- [5] Several proposals have been made as solutions to the SGP. Most have been recognized as not providing valid solutions [Taddeo, Floridi, 2005].

References

- Bozkurt, A., Gilmour, R.F., Sinha, A., Stern, D. and Lal, A. 2009. Insect–Machine Interface Based Neurocybernetics. *Biomedical Engineering, IEEE Transactions on*.56 (2): 1727-1733.
- Floridi, L. 2003. From Data to Semantic Information. *Entropy* 2003, 5: 125-145.
<http://mdpi.muni.cz/entropy/papers/e5020125.pdf>.
- Harnad, S. 1990. The Symbol Grounding Problem. *Physica D* 42: 335-246.
- Menant, C. 2003. Information and Meaning. *Entropy* 2003, 5: 193-204.
<http://mdpi.muni.cz/entropy/papers/e5020193.pdf>.
- Menant, C. 2005, a. Information and Meaning in Life, Humans and Robots. *Proc. Of the 3rd Conference on the Foundation of Information Sciences, Paris*.
- Menant, C. 2005, b. Evolution and Mirror Neurons. An introduction to the Nature of Self-Consciousness. *TSC 2005*. Copenhagen, Denmark. <http://cogprints.org/4533/>.
- Menant, C. 2010. Evolutionary Advantages of Inter-subjectivity and Self-Consciousness through Improvements of Action Programs. *TSC 2010*. Tucson, AZ. <http://cogprints.org/6831/>.
- Menant, C. 2011. Computation on Information, Meaning and Representations. An Evolutionary Approach. in *Information and Computation: Essays on Scientific and Philosophical Understanding of Foundations of Information and Computation*, edited by G. Dodig-Crnkovic and M. Burgin. World Scientific. ISBN-10: 9814295477, pp. 255-286.
- Philpapers, 2013. Nature of Consciousness. <http://philpapers.org/s/nature%20of%20consciousness>.
- Philpapers, 2013. Nature of life. <http://philpapers.org/s/nature%20of%20life>.
- Riofrio, W. 2007. Informational Dynamic Systems: Autonomy, Information, Function. In *Worldviews, Science, and Us: Philosophy and Complexity*, edited by C. Gershenson, D. Aerts, and B. Edmonds. World Scientific, Singapore, pp. 232-249.
- Searle, J. R. 1980. Minds, brains and programs. *Behavioral and Brain Sciences* 3: 417-424.
- Taddeo, M., Floridi, L. 2005. Solving the Symbol Grounding Problem: Critical Review of Fifteen Years of Research. *Journal of Experimental & Theoretical Artificial Intelligence*, 17, (4).
- Turing, A.M. 1950. Computing machinery and intelligence. *Mind*, 59, 433-460.

Warwick, K., Xydias, D., Nasuto, S. J., Becerra, V. M., Hammond, M. W., Downes, J., Marshall, S. and Whalley, B. J. 2010. Controlling a mobile robot with a biological brain. *Defence Science Journal* 60 (1): 5-14. ISSN 0011-748X