The consciousness-model of De Bruijn

Henk Barendregt
Faculty of Science, Radboud Universiteit, Nijmegen
Netherlands Institute of Advanced Study
in the Humanities and Social Sciences, Wassenaar

December 21, 2012

1. Introduction The Dutch mathematician N.G. de Bruijn (9 July 1918-17 February 2012) was a versatile scholar who was still active at an advanced age. He worked on number theory (Sturmian sequences), analysis (asymptotics), discrete mathematics (de Bruijn-Erdos theorem, analysis of Penrose tilings leading to quasi-crystals), and logic (automated proof-checking). He also worked on the model of consciousness [4, 5, 6, 7], that is presented in this paper. The model is partly about the functioning of *memory*. It is conceived from the perspective of mathematical thinking, giving it penetrating elegance and economy.

An *associative memory* (AM) unit is defined as a 'system' that is capable of doing the following. Suppose two elements a, b from a set of 'signals' Σ are presented to the system one after the other within a short interval of time. Then the system 'remembers' that b followed a in the following manner: if the query $a?$ is presented later to the system then it may be able to answer $b!$. The station can fail to give a correct answer or fail to give an answer at all. For example if the memory has exceeded its *capacity* (is filled up) at the time of presenting $\langle a, b\rangle$ it will always fail to answer the query $a?$. The idea to specify memory in terms of this task, called *cued-recall*, has been used early on in experimental psychology, see [8]. The system also has a *reliability*, defined as the chance that the answer is correct. The core of the model of de Bruijn consists of describing an (abstract) way in which a large number ($N = 10^{10}$) of AM units, each with limited capacity and reliability can work together to obtain a compound AM with a large capacity and extremely high reliability. If the compound system satisfies a certain property (the *echo hypothesis*), then the memory capacity can be strengthened by practise.

Using the compound AM De Bruijn is able to define consciousness and subconsciousness. The way this compound memory is constructed has consequences for the functioning of consciousness. Several other aspects of consciousness are also discussed.

2. The memory model The possibility to perform cued-recall correctly for a large number of pairs can arguably be seen as a necessary and sufficient

condition for being considered as a memory.

Desired properties. Suppose we have a number of small AM units. How can one make them collaborate so that the compound has a large capacity and reliability? Preferably fast, robust, and relatively energy economical as well. In order to multiply the capacity by a number k one may start thinking about taking that many small AM units. The pairs $\langle a, b\rangle$ are stored in one of the units, until it is filled up. After that the next pairs are stored in a next unit, etcetera. Then the problem arises that in order to retrieve information one also has to know where it is stored, i.e. its address, and this also uses up memory. Moreover, this solution is not robust, because if the unit where a pair is stored becomes defective, then there will be loss of information.

A 'paradox'. De Bruijn's model provides a memory with large capacity and reliability and that is robust, fast, and relatively economical. It is based on a simple mathematical Lemma that is presented here as a well-known pseudo-paradox.

> *Each time when an adult with normal lung volume has inhaled,*
> *his or her lungs contain at least one molecule oxygen,*
> *that also has been inhaled by Archimede.*

The explanation of this 'paradox' is that Avogadro's number (10^{23}, the number of molecules in a mole) is large, even in comparison to the ratio of the lung-volume of a normal adult female or male (i.e. 3 to 6 litres) and all the air on Earth.

Constructing a large memory. The model of a large memory consists of an operating system that coordinates a large number of small memory units in the following way. The small units are made to be active (ON) at random, during a short time interval, waiting for information to store or a query to answer. For most of the time they are not active (OFF). De Bruijn takes the avarage 0.5 sec (ON) against 4 hours (OFF), a ratio in the order of 1 to 10^4, as typical values. During the learning phase the operating system sends a pair $\langle a, b\rangle$ to all units. Only those units that are switched ON will register the information. At a later point in time the query a? could arrise, and is sent to all units. If there are $N = 10^{10}$ units, then at every single moment there will be an average 10^6 units switched ON. This means that at the moment the query is made, the expectation is that there are 100 units switched ON, that were also ON during the learning phase $\langle a, b\rangle$. [The argumentation is the same as for understanding the pseudo-paradox.] From 100 answers, possibly incorrect, the right answer can be constructed by the operating system based on a majority vote. [For this it must be assumed that the possible mistakes are not made in a systematic way.] The capacity of the large AM unit is in the order of $\sqrt{N}$ times the minimum capacity of the small AM units. In case these all have a reliability of only 0.5, then the reliability of the large memory is $1 - e^{-20}$. The large AM is fast (provided that there is a fast way to reach all cells), robust[, and relatively energy-efficiënt]. [Yet every moment 10^6 units are required to be ON, explaining why the brain consumes a lot of energy. The fact that not all 10^{10} are ON explains the relatieve

efficiency.] For the derivations see [6].

The small AM units. Their functioning was imagined by De Bruijn as follows. The information carrying elements $a, b, \ldots \in \Sigma$ correspond to chain-like molecules $A, B, \ldots$. When a pair $\langle a, b\rangle$ is presented to the small system, then molecules A followed by B are released in the unit, thought of as a cell, after which an enzyme forms the molecule $A * B$. Immediately after that A and B disappear. When later the quary $a?$ is being asked, then a molecule $A?$ is being produced. Then there is a search for the molecule of the form $A*B$, and from it, by means of another enzyme, the molecule $B!$ is being produced, which is a coding of the answer $b!$. The nature of the information carriers and the translation mechanism to molecules is not discussed by De Bruijn.

If one neuron can serve as the small AM unit, then for the large AM there are $N > 10^{10}$ of these available. As 'AM unit' is an abstract concept, it also may (have to) be interpreted as a neural net, of which the brain contains many more.

Operating system. De Bruijn also had ideas how the operating system could function. He imagined a tree of branching neurons, in which each neuron dynamically makes contact with a next neuron. [De Bruijn imagined moving synapses. Because of the necessary speed, this possibly can be better realized by synapses to a relay, where a pulse may be directed to the next neuron or be stopped.] Those neurons that are at moment t in connection with the root of the tree can be seen as the set $A(t)$, ready for storage and retrieval of information. The simple statistical analysis by which the capacity and reliability of the large AM was computed above will have to be modified, because the units are no longer independent: the distance to the root of the tree now plays a role. [This may play a role to improve an unrealistic aspect of the model of the large AM, mentioned in [7]: in the human brain there are specialized areas. Moreover the brain is less reliable than $1 - e^{-20}$.]

Strengthening memory. [Sometimes memory cannot retrieve which association was stored. This may be caused by unfavorable parameters in the model so that in $A(t)$ the answer to the query is not stored. Alternatively the tree-model of $A(t)$ may have an unfavorable effect on the reachability of the units that do know the answer.

The following hypothesis is being proposed.] During information retrieval following the query $a?$ the answer $b!$ comes. The 'echo-hypothesis' in [4, 5] states that after successful retrieval the pair $\langle a, b\rangle$ is being broadcasted once more, thereby inducing more small units to remember the association. In this way the compound memory can improve its capacity by *repetition.*

3. Consciousness According to De Bruijn several aspects of consciousness can be connected to his model of memory. At moment t there is a set of associative memory cells $A(t)$ that is active (ON). De Bruijn calls it the *'roaming random set'.* In the view of De Bruijn $A(t)$ is not a fixed buffer in which information comes and goes. The memory units are determined by a moving

window, comparable with one that is used in a text-editor on a computer.

Input $a \in \Sigma$ comes at first from the outside world. This becomes later a query $a?$ posed to the memory; at moment t it may give an answer $b!$ from $A(t)$. [Many of these answers have to do with inner actions (thinking) or outer actions (moving or speaking).] All these queries $a?$ and answers $b!$ form the potential components of conscious experience. However, not all information coming from the small AM units in $A(t)$ is conscious. De Bruijn calls it '*subconscious*'. [At present it would be called *'pre-conscious'* [13].] For consciousness something more is needed.

Reflection. To each $A(t)$, the set of Am units that is ON at moment t, there is the associated set $A^{\sim}(t)$ of signals, i.e. elements of Σ, that are being processed in $A(t)$ (either as query or answer). These are possible input for a next roaming random set. In order to go from the memory model to a model of consciousness a special operation is assumed to be performed: *knowing*. A signal c can be 'known', and then a signal $\underline{c}$ will result, that is the 'knowing of c'. This is called 'thinking about c' by De Bruijn. [Presently one uses *meta-awareness* [14].] This form of reprocessing of information is called 'reflection'. [The advantage of reflection is the following. Suppose that the pairs $\langle a_1, a_2\rangle, \langle a_2, a_3\rangle, \langle a_3, a_4\rangle, \ldots$ have been recorded. Then the *scenario* $a_1 - a_2 - a_3 - a_4 - \ldots$ results from this chain of associations. In case one can work with meta-awareness, then mentioned scenario can be considered, without the need to 'run it', becoming $\underline{a}_1 - \underline{a}_2 - \underline{a}_3 - \underline{a}_4 - \ldots$. That is, one processes the signals without automatically reacting. One could say about the scenario, as a mathematician, to run it 'on paper', or as a computer scientist 'as code', respectively.] The set $R^{\sim}(t)$ as defined as the subset of $A^{\sim}(t)$ consisting of the reflective elements '$\underline{a}$'. These deal with the elements of what is called by De Bruijn consciousness.

Thinking. For the development of (mathematical) thinking the act of *computing* is less important than that of *pattern matching*. The latter is comparable to solving a jigsaw puzzle: one tries haphazardly, and if a few pieces do fit, one has obtained an isle using which one can continue to build. In collaboration with reflection several levels of abstracion may result.

Personality. The strange feeling that our personality is a spectator of the mind-body theater may be related to reflection. The stance of solipsism is seen by De Bruijn as neither true nor false, but mainly as a sterile attitude [5].

The memory model is sufficiently flexible that several operating systems can use the large memory simultaneously. [This is reminiscent of the phenomenon of 'multiple personality'. The model could explain the possibility of the presence or absence of common knowldege of these personalities.]

The hard problem. The attitude of De Bruijn towards solving the body-mind problem was modest. He didn't have the pretention to have solved the *hard problem*, [2], asking for an explanation of 'experience'. "It seems that no theory whatsever can solve this difficulty", according to De Bruijn in [5].

Consequences. One of the consequences of the model of consciousness is that

it is *transient*. The roaming random set only exists in the order of a second. After the disappearing of $A^{\sim}(t)$ and $R^{\sim}(t)$ new versions appear. The contents of consciousness can be made stable by repeating. This is the mechanism of *concentration* [5].

Deciding whether a phenomenon appears in consciousness coming from the external world or from a phantasy or dream cannot be determined by the mechanism, but only on be based on criteria of consistency of the content.

Free will. The position of De Bruijn was that people have no free will, but they do experience it. This doesn't imply that determinism at lower levels does or doesn't exist. The effect of a large number of deterministic subprocesses may be considered at the next level as a *random* process. Non-determinism at quantum level may be considered as deterministic at the next level, at the next level again as random, one level higher as satisfying a strict logic and finally at the highest level be interpreted as a creative process with serendipity [5].

Other models. De Bruijn welcomed other models, notably to see to what extent they contradict each other and they agree. [Modern theories on consciousness in Fig. 1, similar to the one of De Bruijn, all employ two layers: $C_1 \supseteq C_2$. The layer C_1 can be considered as *'basis-consciousness'* and the part C_2 as a more specific form. In the model of De Bruijn these are $A^{\sim}(t)$ and $R^{\sim}(t)$. There is a difference in name and the way in which the second layer is being determined. See Fig. 1. In all models there is also a notion of *unconscious*. This is implicit in that of De Bruijn, where one can define this notion as the state of all small memory units that are OFF, and explicitly in the other models.

It seems attractive to explain the difference between *attention* and *reflection* as follows. Attention is a way to choose among the elements of C_1. Indeed, the three authors that employ attention to differentiate between C_1 and C_2 call it *selective attention*. This selection does not necessarily happen consciously, most of the time it may even be completely unconscious. The notion 'reflection' may be seen as a conscious variant of attention: selective attention that is conscious in the sense of C_1 (the left column of Fig. 1). Carrying on along these lines one could interpret *mindfulness* [12] as selective attention that is conscious according to C_2 (the right column of Fig. 1). These considerations fall outside the scope of this paper, but show that the model of De Bruijn may play a role in contemporary discussions.]

4. Reception of the model The first ideas about the functioning of human consciousness were presented by De Bruijn an a meeting of the Dutch Royal Academy in 1974 in a talk titled: "*Mathematical models of the living brain*" [3]. The paper did not yet contain the notion 'roaming random set'. During the discussion there was wave of criticism: "This talk has nothing to do with the brain!", was the reproach. The title of the talk was chosen indeed in a suboptimal way. The ideas were about human consciousness, in particular on memory. In later papers [4, 5, 6, 7], appearing after about twenty years, the methodology of a methematical was explained more and more explicitly. "Because the body-mind problem [how does consciousness function on the basis of what is known

	from $C_1 - C_2$	via	to C_2
De Bruijn	*subconsciousness*	R	*consciousness*
Edelman [9]	*consciousness*	R	*reflective consciousness*
Hobson [10]	*protoconsciousness*	R	*higher-order consciousness*
Dehaene [13]	*preconsciousness*	A	*consciousness*
Block [1]/ Lamme [11]	*phenomenal consciousness*	A	*access consciousness*

Figuur 1: Models of consciousness and terminology. R: *reflection*, A: *attention*.

about the body?] is not solved—neither by (molecular) biology, nor by (cognitive) psychology, and also not by the neurosciences—this question appears in nomansland. It is here that a mathematician may contribute, knowing what it means to have a model: it doesn't necessarily imply truth and language and metalanguage have to be distinguished. By being familiar with mathematical reasoning, computer architecture and probability one may make a contribution, even if the nomansland is a mine-field because the mathematician may have too little knowledge of mentioned disciplines." Said De Bruijn in [7]: 14. Apology [citation not literally].

From the fact that there are five different publications over the model of consciousness, and from personal discussions with De Bruijn, it became clear that he found it important that his model became more known.

Postscript The text between square brackets are added for clarification by the author. I thank Fabio Giommi, Hans Maassen, Antonino Raffone and Marieke van Vugt for useful discussions and information. This work was supported by a Grant from The Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS).

Referenties

[1] N. Block. Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30(5-6):481–499, 2007. Discussion 499-548.

[2] D. Chalmers. *The Conscious Mind*. Oxford University Press, 1996.

[3] N.G. de Bruijn. Wiskundige modellen voor het levende brein. *Verslag van de Gewone Vergadering Afd. Natuurkunde*, 83:156–164, 1974.

[4] N.G. de Bruijn. A model for information processing in human memory and consciousness. *Nieuw Archief Wiskunde*, 4(12):35–48, 1994.

[5] N.G. de Bruijn. Can people think? *Journal of Consciousness Studies*, 3:425–447, 1996.

[6] N.G. de Bruijn. A model for associative memory, a basis for thinking and consciousness. In P. van Emde Boas J. Wiedermann and M. Nielsen, editors, *Automata, Languages and Programming,* Proceedings of the 26th International Colloquium, ICALP '99, Prague, volume 1644 of *Lecture Notes in Computer Science*, pages 74–89. Springer, 1999.

[7] N.G. de Bruijn. A mathematical model for biological memory and consciousness. In Fairouz D. Kamareddine, editor, *Thirty Five Years of Automating Mathematics*, pages 9–23. Kluwer Academic Publishers, 2003.

[8] H. Ebbinghaus. *On memory: A contribution to experimental psychology.* Teachers College, Columbia University, New York, 1885/1913.

[9] G.M. Edelman. *The Remembered Present: A Biological Theory of Consciousness.* Basic Books, New York, 1990.

[10] J.A. Hobson. REM sleep and dreaming: towards a theory of protoconsciousness. *Nature Reviews Neuroscience*, 10(11):803–862, 2009.

[11] V.A.F. Lamme. Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7(1):12–18, 2003.

[12] A. Raffone and N. Srinivasan. An adaptive workspace hypothesis about the neural correlates of consciousness: insights from neuroscience and meditation studies. In N. Srinivasan, editor, *Progress in Brain Research: Attention*, pages 161–180. Elsevier, 2009.

[13] Dehaene S., Changeux J.P., Naccache L., Sackur J., and Sergent C. Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn Sci.*, 10(5):204–211, 2006.

[14] J.W. Schooler. Re-representing consciousness: Dissociations between consciousness and meta-consciousness. *Trends in Cognitive Science*, 6:339–344, 2002.