Axiomatizing Consciousness with applications

Henk Barendregt
Faculty of Science
Radboud University
Nijmegen, The Netherlands

Antonino Raffone
Faculty of Psychology
Sapienza University
Rome, Italy

‘Axiomatization’ or modeling, as approximate structure

‘Applications’ mainly to suffering and its release

Inspiration comes from

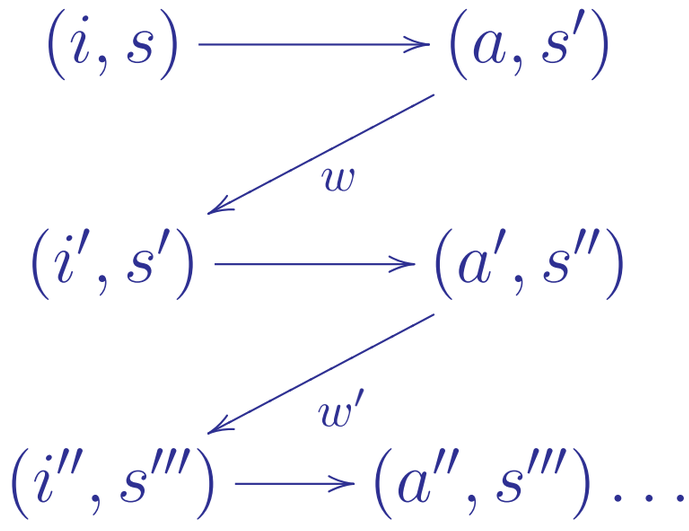
• Turing machines (including type B machines) • Buddhist psychology • Meditation experience • Emperical evidence	} discreteness
• Friston’s Free Energy Principle	life is risky
• N.G. de Bruijn’s model of consciousness	associative memory

Central notion: (mind-)state s determining *actions, experience, physiology*

[unwholesome states may bring us in contact with police, psychiatry, physicians
wholesome ones increase chances for creativity, flow and health]

states appear in time $\{s_t\}_{t\in\mathbb{T}}$ and conscious time is discrete $\mathbb{T}=\mathbb{Z}$

$i \in \Sigma$ (input), $s \in Q$ (state), $a \in A$ (action), $w \in W$ (world)
an action a can be a move $\{L, R\}$ or $\{\text{write}(i)\}_{i \in \Sigma}$



Equivalent rendering

$$(i, s, a) \xrightarrow{w} (i', s', a') \xrightarrow{w'} (i'', s'', a'') \xrightarrow{w''} (i''', s''', a''') \ \ldots$$

$$(i, s, a; w) \longrightarrow (i', s', a'; w') \longrightarrow (i'', s'', a''; w'') \longrightarrow (i''', s''', a'''; w''') \ \ldots$$

Add sensors and actuators to the TM

Generalize Σ to $\mathcal{I}$, Q to $\mathcal{S}$, A to $\mathcal{A}$; transitions by a NN

Now $i \in \mathcal{I}$ has a zillion possibilities, similarly for $a \in \mathcal{A}$

This is the Hybrid Turing Machine Model in

Zylberberg-Dehaene [2011], Barendregt-Raffone [2013]

To model transitions $(i, s, a; w) \longrightarrow (i', s', a'; w') \longrightarrow (i'', s'', a''; w'') \ldots$ we interpret states as actors:

$$\mathcal{S} = (\mathcal{I} \times \mathcal{A} \times \mathcal{W}) \to (\mathcal{I} \times \mathcal{S} \times \mathcal{A} \times \mathcal{W})$$

$$\mathsf{next}(i, s, a; w) = s(i, a; w) = (i', s', a'; w'), \text{ etcetera}$$

This if the world w is as obedient as a TM tape

To model *resistence*, consider $\mathcal{W}^*$, the agent's intentions, with a stochastic $\mathbf{w} : \mathcal{W}^* \times \mathcal{W} \to \mathcal{W}$ (prediction correction)

Intended w^* in $(\ldots, w^*) = s(\ldots, w)$ gets 'corrected'

Friston considers things from the other side and speaks about prediction error

Reality view: focus on the result, the corrected intention;

Friston's view: focus on the not accomplished intention, an error

Now we model things as follows

$$\mathcal{S} = (\mathcal{I} \times \mathcal{A} \times \mathcal{W}) \to (\mathcal{I} \times \mathcal{S} \times \mathcal{A} \times \mathcal{W}^*)$$

$$\begin{aligned} \mathsf{next}(i, a; w) &= (i', s', a'; w'), \text{ with } w' = \mathbf{w}(w^*, w) \text{ and} \\ &\quad (i', s', a'; w^*) = s(i, a; w) \end{aligned}$$

Sensorial input comes in parallel, but one can focus

$i = \langle \{i_1, \ldots, i_n\}, F \rangle$ with $F \subseteq \{i_1, \ldots, i_n\}$ (attention)

Also for actions (empirical data for animals exists)

$a = \langle \{a_1, \ldots, a_m\}, G \rangle$ with $G \subseteq \{a_1, \ldots, a_m\}$

An action moves the world or attention (like in a TM!)

A state consists of coalition of collaborating substates

$$s = q_1 | q_2 | \cdots | q_k$$

[A state can be approximated by a large set of parameters

and a substate is a subset of these (wheather: a snow storm)]

An important substate is q_{f} *feeling-tone*: pleasant, unpleasant, neutral

Axioms for consciousness

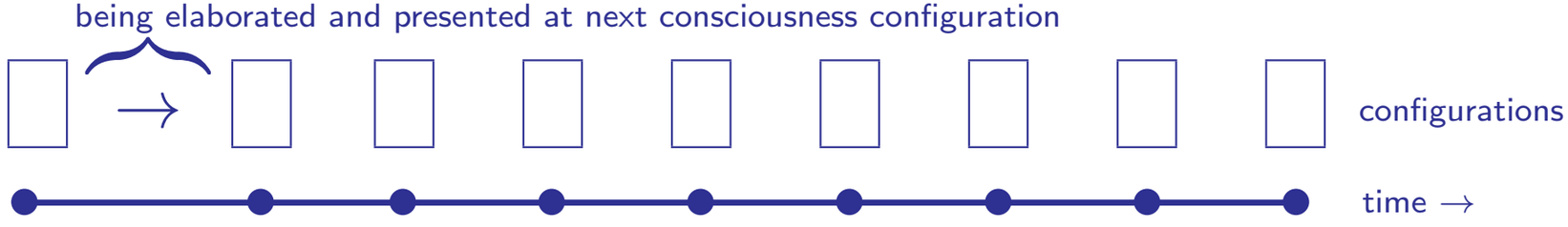
1. $\tilde{\mathbf{c}} = \{\mathbf{c}_t\}_{t\in\mathbb{T}}$ consciousness as time-stream of configurations

2. $\mathbf{c} = (i, s, a)$ each $\mathbf{c}$ consists of an object, state and action

3. $i = \langle\{i_1, \ldots, i_n\}; F\rangle, \quad a = \langle\{a_1, \ldots, a_m\}; G\rangle, \quad s = q_1|\cdots|q_k$

4. $\mathbb{T} = \mathbb{Z},$ discrete time with $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$
   'flexibly embedded' into real time $\mathbb{R}$

5. A *pre-conscious* moment is *not a snapshot* but an elaboration, like spatial/temporal interpolation, of what happened in the preceding real time interval



7. Evidence and consequences

Partial evidence

- Lehmann et al [20] — Atoms of the mind
- VanRullen, Koch [26] — Continuous Wagon-wheel illusion

The model allows for spatial/temporal interpollation

Consequences

- Brouwer [4] — Time perception: difference between $\mathbf{c}_t$ and $\mathbf{c}_{t-1}$ the latter being still somewhat accessible
- von Neumann — "How to explain precision under biological noise?"
  Zylberberg, Dehaene et al [30]: by discretization
- Wertheimer [29] — The Φ-phenomenon
- Gestalt psychologists — The illusion of the sensory mosaic
- Sperling [23] — Unconscious attention
- Slagter, Lutz et al. [31] — Variable attentional blink

The stream of consciousness so far is

$c_1, c_2, c_3, \cdots$ with $c_t = (i_t, s_t, a_t)$

Focussing on components we get

$i_1, i_2, i_3, \cdots$	*stream of input*
$s_1, s_2, s_3, \cdots$	*stream of states*
$a_1, a_2, a_3, \cdots$	*stream of actions*

The i may be experiential i_{s}, from the physical senses or from the mind i_{m} and often the stream is rather mental

$$i_{\mathsf{s}}, i_{\mathsf{m}}, i_{\mathsf{m}}, i_{\mathsf{m}}, \cdots$$

the *narrative self*. This may be unwholesome if it is coupled with states containing negative feeling-tone and aversion

Basic mechanism of associative memory: *cued recall*

When an object i_1 is followed by i_2 a couple of times,

then one remembers this and presenting i_1 will produce i_2 from memory

There is a mental *storing* factor q_s such that a few

$$i_1(s|q_s)a_1 \to i_2(s|q_s)a_2,$$

i_1 is followed by i_2 creates a mental *recall* factor $q_{r[i_1,i_2]}$ such that in the recall phase

$$i_1(s|q_{r[i_1,i_2]})a \to i_2sa$$

i.e. the object i_2 follows i_1. Can create thought chains:

$$i_m \to i'_m \to i''_m \to \cdots$$

10. Associative memory

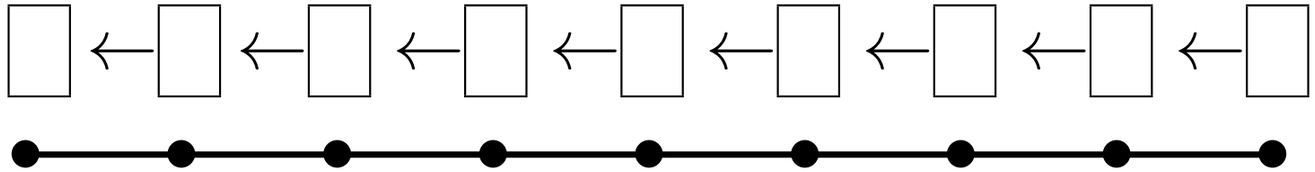
Units of cued recall (CR) have limited capacity and reliability. Many such units can be combined to form powerful associative memory (AM), enabling it to remember a text of 500 pages with very high reliability

Theorem (NG de Bruijn [3]) *Suppose that each unit of CR has a capacity of storing* m *pairs, with a reliability of* 0.5 (50%) *to be correct. Suppose there are* $N = 10^{10}$ *such units, then it is possible to construct a compound system for AM, with capacity* $\sqrt{N}.m$ *and reliability* $1 - e^{-20}$.

Proof. Sketch. The 10^{10} small units of CR are made to be On/Off at random in a ratio $1/10^4$ (0.5s On vs 4h Off). Then by simple probability at every moment 10^6 units are On. During learning phase these store the $\langle i_1, i_2 \rangle$. At recall time there are also 10^6 On (a different set). Of these, 10^2 also had been On during storing time. A majority vote provides the right association. ■

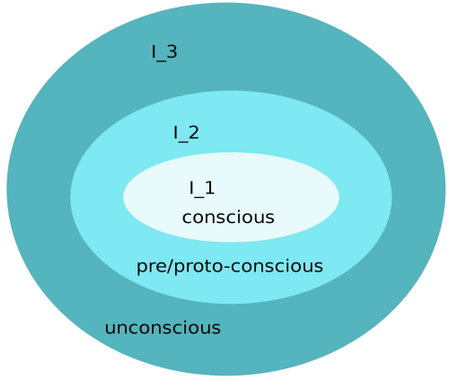
Attention was built in the model as $i = \langle \{i_1, \ldots, i_n\}; F \rangle$

Mindfulness (reflection) input from precedent configuration not from the world: can be i, q_{f}, s, a (foundations of mindfulness)



Block [3], Lamme [18], Dehaene et al. [8] claim attention creates full consciousness

de Bruijn [5], Edelman [10] and Hobson [14] think one needs more: reflection



Distinguishing an extra level, using attention and reflection differentially, one gets pre-consciousness vs (full-)consciousness deserves to be explored

Worldly suffering

coming from the prediction correction causing unpleasant feelings (pain) and worse aversion or greed to cover it up. There may be a way out by improving the situation

Existential suffering ($\rightarrow$ existential fear)

coming from seeing that we are machine-like with discrete consciousness that can be chaotically jumpy rolling on 'without-self'

In both cases one wants to be in control, only partially possible in the first case. We try to protect ourselves by 'addictive' states (side-effects)

Notably existential suffering brings us in strange attractors in the dynamic system of consciousness streams hitting frequently unpleasant feeling. Attempts to cover it up, even if partially successful, make us rigid

Skillfully applying mindfulness

$(i, $  $, a!) \rightsquigarrow (i + $ 'angry',  $, a)$ *naming*, or better *noting*:

$(i, $  $, a!) \rightsquigarrow (i + $ '  ',  $, a)$ ($\sim$ Turing completeness)

Then 'angry' can be avoided from focus or be reinterpreted

Disentangling $(i, s, a) = (\langle\{i_1, \ldots, i_n\}, F\rangle, q_1|q_2|\cdots|q_k, \langle\{a_1, \ldots, a_m\}, G\rangle)$

it is possible that mind-states and behavior become more flexible

Given strange attractor $(i_1, s_1, a_1) \to \cdots \to (i_n, s_n, a_n) \sim (i_1, s_1, a_1)$

developing concentration, fixing the i and the a yields

$(i, s_1, a) \to \cdots \to (i, s_n, a) \sim (i, s_1, a)$

Then clearly seeing the loop, a reset happens: system leaves attractor

From narrative self, subject to emptiness ($\to$ existential fear), to experiential self

14. Research theme

Study comprehensively the putative discrete nature of consciousness and the consequences of perceiving this:

fear, delusion, disenchantment [32]

Study the mechanism of overcoming the fear by

1. the capacity of mindful reification
2. Insight that consciousness is a dynamic process

and refining consciousness with improved concentration, friendly attention, and mindfulness

If correct, this may help understanding clinical conditions

[0] Anuruddha. A comprehensive manual of Abhidhamma (Abhidhammattha Sangaha). Translation by Mahathera Nārada, edited by Bikkhu Bodhi, U Rewata Dhamma. *Abidhammattha Sangaha*, Buddhist Publication Society, Kandy, 1993.

[1] Baars. The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences*, 2002, 6(1), 47-52.

[2] Barendregt, Raffone. Conscious cognition as a discrete, deterministic, and universal Turing machine process. Alan Turing - His Work and Impact. Eds. Cooper and van Leeuwen, Elsevier, 2013, 92-97.

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

[4] Brouwer. Consciousness, Philosophy, and Mathematics, *J of Symbolic Logic*, 1949, 14(2), 132-133.

[5] de Bruijn. A mathematical model for biological memory and consciousness, in: Kamareddine (ed.), *Thirty Five Years of Automating Mathematics*, Kluwer Academic Publishers, 2003, 9-23.

[6] Chalmers. Facing up to the problem of consciousness. *J of Consciousness Studies*, 1995, 2(3) 200-219.

[7] Creswell, Lindsay. How does mindfulness training affect health? A mindfulness-stress-buffering account. Current Directions in Psychological Science, 2014. doi:10.1177/0963721414547415.

[8] Dehaene, Changeux, Naccache, Sackur and Sergent. Conscious, preconscious, and subliminal processing: a testable taxonomy, *Trends Cogn Sci.* 10(5), 2006, 204-211.

[9] Ebbinghaus. *On Memory*. Dover, 1885/1964.

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

[11] Eigen. *The Psychotic Core*. Karnac Books, 2005.

[12] Fuxe, Agnati. *Volume Transmission Revisited*, Elsevier, 2000.

[13] Fuxe, Agnati. Cell-Cell Communication through the Extracellular Space.

Squire (ed.) *Encyclopedia of Neuroscience*, Academic Press, 2, 2009, 655-664.

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

[15] Jacobs, Epel, Lin, Blackburn, Wolkowitz, Bridwell. Intensive meditation training, immune cell telomerase activity, and psychological mediators. Psychoneuroendocrinology, 36, 2012, 664-681.

[16] Kelley, Lambert. Mindfulness as a potential means of attenuating anger and aggression for prospective criminal justice professionals. Mindfulness, 3, 2012, 261-274.

[17] Khoury, Lecomte, Fortin, Masse, Therien, Bouchard, Chapleau, Paquin, Hofmann. Mindfulness-based therapy: A comprehensive meta-analysis. Clinical Psychology Review, 33, 2013, 763-771.

[18] Lamme. Why visual attention and awareness are different, *Trends in Cognitive Sciences*, 7(1), 2003, 12-18.

[19] Levine. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 64, 1983, 354-361.

[20] RD Pascual-Marqui, CM Michel, D Lehmann. Segmentation of Brain Electrical Activity into Microstates: Model estimation and Validation. *IEEE transactions on Biomedical Engineering*, 42, 1995, 658-665.

[21] Salzman, Fusi. Emotion, Cognition, and Mental State Representation in Amygdala and Prefrontal Cortex. *Annu. Rev. Neurosci.* 2010(33), 173-202.

[22] Skarda, Freeman. How brains make chaos in order to make sense of the world, *Behavioral and Brain Sciences*, 1987, 10, 161-195.

[23] Sperling. The information available in brief visual presentations, *Psychological Monographs: General and Applied*, 74(11), 1-29.

[24] Turing. On computable numbers, with an Application to the Entscheidungsproblem, *Proceedings of the London Mathematical Society*, 2(42), 1937, 230-265.

[25] Turing. Intelligent machinery, report for National Physical Laboratory, published in Machine Intelligence 7, B. Meltzer and D. Michie (eds.) 1969; also in: Ince (ed). Collected works of AM Turing — Vol 1: Mechanical Intelligence. Elsevier Science Publishers, 1992, 3-23.

[26] VanRullen, Koch. Is perception discrete or continuous? *Trends Cognit Science*, 7, 2003, 207-213.

[27] Varela. Neurophenomenology: a methodological remedy for the hard problem, *Journal of Consciousness Studies*, 3(4), 1996, 330-349.

[28] Veening, Barendregt. The regulation of brain states by neuroactive substances distributed via the cerebrospinal fluid. A review. *Cerebrospinal Fluid Research*, 2010, 7(1).

[29] Wertheimer. Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*. 1912, 61(1), 161-265.

[30] Zylberberg, Dehaene, Roelfsema, Sigman. The human Turing machine: a neural framework for mental programs. *Trends in Cognitive Sciences*, 2011, 15(7), 293-300.

[31] HA Slagter, A Lutz et al Mental Training Affects Distribution of Limited Brain Resources, 2007, *Plos Biology*, <doi.org/10.1371/journal.pbio.0050138>.

[32] Buddhaghosa, The Path of Purification (Visuddhimagga), 2011, Bhikkhu Nyanamoli (trans.), Buddhist Publication Society, Kandy.