frontiers in Psychology

OPINION
published: 30 July 2020
doi: 10.3389/fpsyg.2020.01675

Consciousness, Free Energy and Cognitive Algorithms

Thomas Rabeyron[1,2]* *and Alain Finkel*[3]

[1] Psychology Department, Université de Lorraine, Nancy, France, [2] Psychology Department, University of Edinburgh, Edinburgh, United Kingdom, [3] École normale supérieure Paris-Saclay, Cachan, France

Keywords: consciousness, entropy, free energy, algorithms, subjectivicty

CONSCIOUSNESS STUDIES: FROM THE BAYESIAN BRAIN TO THE FIELD OF CONSCIOUSNESS

Different theoretical approaches have tried to model consciousness and subjective experience, from phenomenology (Husserl, 1913), cognitive psychology and neuroscience (Baars, 2005; Dehaene, 2014), artificial intelligence and cybernetics (Baars and Franklin, 2009; Rudrauf and Debban, 2018), statistical physics and probabilistic models (Solms and Friston, 2018) to mathematics of relationships (Ehresmann and Vanbremeersch, 2009). In this regard, even if the neurobiological functioning of the brain is different from the symbolic processing of a computer (Varela et al., 1991), it might be relevant to conceptualize psychological activity as a Turing machine. For example, Dehaene (2014) assumes that this type of machine offers "a fairly reasonable model of the operations that our brain is capable of performing under the control of consciousness" (p. 151) and points out that "the conscious brain (...) functions like a human Turing machine that allows us to mentally execute any algorithm[1]. Its calculations are very slow, because each intermediate result must be stored in working memory before being sent to the next step—but its computing power is impressive" (p. 150). From this point of view, we would like to suggest in this paper how we could rely on fundamental tools used in computer sciences such as computability theory, algorithmic and finite automata (Pin, 2006; Wolper, 2006) in order to improve our understanding of consciousness.

Among current theories of consciousness, one of the most promising has been developed during the last 10 years by Karl Friston (2009) which states that the brain constructs a predictive representation of its environment that infers the probable causes of sensory stimuli. This representation, or simulation, would lead, or could even be equal, to consciousness (Solms, 2013). This predictive model cannot be right all the time and sometimes there must be a "gap" between the probabilistic representation of the world produced by the brain and the actual perceptual data coming from the environment. It engenders an increase of entropy and free energy in the brain, which would induce subjective feelings of surprise (Friston, 2009; Carhart-Harris and Friston, 2010). Thus, to reduce entropy and free energy, the brain improves progressively its Bayesian probabilistic model of the potential cause of its sensations based on previous assumptions.

Continuing the Bayesian brain hypothesis and Friston's work, Rudrauf and his colleagues (Rudrauf and Debban, 2018; Williford et al., 2018) recently introduced the "Projective Consciousness Model" (PCM) which is a projective geometrical model of the perspectival phenomenological structure of the field of consciousness[2]. The PCM accounts for "the states of the agent's body in its relations to the world and to others by being constantly quantified by the processes of active inference" (Rudrauf and Debban, 2018, p. 161). Its main function is to

[1] We do not assume that the brain is a computer—as proposed in the symbolic approach whose weaknesses opened the way to connectionism (Varela et al., 1991)—but we suppose that we can analyze part of its functioning using original tools borrowed from the field of computer studies.

[2] From a more philosophical point of view concerning how the PCM could be articulated with the hard problem of consciousness (Chalmers, 1996) and the infinite regress argument (Dennett, 1991), see Williford et al. (2018), especially on p. 10–14.

OPEN ACCESS

Edited by:
Antonino Raffone, Sapienza University of Rome, Italy

Reviewed by:
Marco Mirolli, National Research Council (CNR), Italy

***Correspondence:**
Thomas Rabeyron
thomas.rabeyron@gmail.com

Specialty section:
This article was submitted to Consciousness Research, a section of the journal Frontiers in Psychology

Received: *10 April 2020*
Accepted: *19 June 2020*
Published: *30 July 2020*

Citation:
Rabeyron T and Finkel A (2020) Consciousness, Free Energy and Cognitive Algorithms. Front. Psychol. 11:1675. doi: 10.3389/fpsyg.2020.01675

reduce free energy and "realize a projective geometrical rendering engine embedded in a general active inference engine, which in turn is presided over by a global free energy minimization algorithm" (Williford et al., 2018, p. 9). The PCM is more precisely composed of a (1) "World Model," mainly unconscious, which stores in memory all the agent's prior beliefs and generative models (2) the "Field of Consciousness" (FoC) which is an explicit model of subjective and conscious experience which takes the form of a simulation in three-dimensions. The FoC represents the sensory perceptions and scenes imagined at any given moment with a specific point of view and can be studied thanks to a domain of mathematics called projective geometry.

MODELING OF THE SUBJECTIVE EXPERIENCE USING COGNITIVE ANALYSIS

Complementary to these computational approaches and "third person point of view" of brain functioning, methods inspired by phenomenology have been developed—explicitation interview (Maurel, 2009; Vermersch, 2012) or micro-phenomenology (Petitmengin and Bitbol, 2009; Bitbol and Petitmengin, 2017)—in order to improve our understanding of subjective experience from the "first-person point of view[3]." One of these neurophenomenological approaches called Cognitive Analysis (CA) has been recently developed by Finkel (1992, 2017) and Tellier and Finkel (1995). CA uses specific interview techniques and modeling tools aimed at describing subjective experience (Finkel, 2017). It also differs from other neurophenomenological approaches by relying on research conducted on mental representations (Kosslyn and Koenig, 1995; Pearson and Kosslyn, 2013) and by using tools from fundamental computer sciences (finite automata and algorithms) following in particular the work of Fodor (1975, 1979, 1983).

CA permits a precise description of the succession of representations used by an individual in order to get closer to his subjective experience (Finkel, 2017). Mental activity is broken done more precisely into three main types of mental objects: sensations (visual, auditive, or kinesthetic), emotions (primary and secondary) and symbolic (verbal language). These mental objects "appear" within the attentional buffer which is itself connected to a long-term information storage system. The subjective experience will also rely on the attentional processes that can be focalized on the internal or the external world. The stream of consciousness can then be conceptualized as a cognitive algorithmic sequence, i.e., a finite sequence of internal and external states and actions (Finkel and Tellier, 1996). Subjective experiences of variable complexity can be analyzed in this way, whether they concern a simple phenomenological experience (e.g., recalling a lived scene), a simple cognitive task (e.g., an addition) or a more complex phenomenological experience (e.g., an Out of Body Experience, see Rabeyron and Caussie, 2016). We thus obtain an algorithm which is a synthetic representation of the successive mental states and the actions carried out during each of these states.

The detailed analysis of a sequence lasting a few seconds sometimes require an interview lasting several hours (Rabeyron, 2020), underlying the incredible density of mental representations and operations that characterize conscious and subjective experience. These cognitive algorithms represent an extremely fast succession of representations concerning the internal and the external world composed of sensations, emotions and words. From this point of view, it is interesting to note that the degree of self-reflexivity of the subject is often limited toward his own mental processes. This is the consequence of the speed with which the representations follow one another and the fact that the subject usually pays limited attention to them during ordinary states of consciousness[4]. It is also possible to compare several interviews with the same individual in order to identify recurring patterns and obtain specific cognitive styles (Tellier and Finkel, 1995). This highlights that the same individual usually uses a finite number of algorithms to handle a wide variety of tasks and situations.

CONSCIOUSNESS, COGNITIVE ALGORITHMS, AND THE REDUCTION OF FREE ENERGY

We are now going to describe how PCM and CA could be associated in order to improve our understanding of consciousness. In this regard, we need first to recall that for Williford et al. (2018) "the PCM combines a model of cognitive and affective dynamics based on variational and expected Free Energy (FE) minimization with a model of perspective taking [or a "Field of Consciousness" (FoC) embedding a point of view] based on 3D projective geometry"(p. 2). From this point of view, we can consider that the brain produces a "virtual reality" whose fundamental function is to help the individual to interact with its environment in order to reduce entropy and free energy (Hobson et al., 2014). What is described by Williford et al. (2018) can be conceived as the biological "hardware" necessary to create this virtual reality model—as well as the geometric proprieties of this three-dimensional space—but not the "software" that is used by consciousness to reduce entropy. CA may describe conscious experience in a sufficient detailed manner to determine these "mental softwares" or "mental programs." We thus hypothesize that the brain integrates and develops specifics cognitive algorithms in order to reduce free energy. A synthesis of these different elements is proposed in **Figure 1**.

These algorithms can concern all the behaviors and mental functioning. For example, experts in any field (a scientist, a football player, a pilot, etc.) will be rarely surprised by new events

[3]About the reliability and the validity of such a neurophenomenological approach, see in particular the arguments proposed by Petitmengin (2009).

[4]The degree of reflexivity can be limited during a task but the neurophenomenological approach makes it possible to access afterwards what happened during the task. The participants are usually very surprised to discover what happened in their mind without noticing it during the task itself, because the stream of consciousness is particularly rich and fast. In other words, they were conscious but they were not "thinking about their thinking" during the task itself.

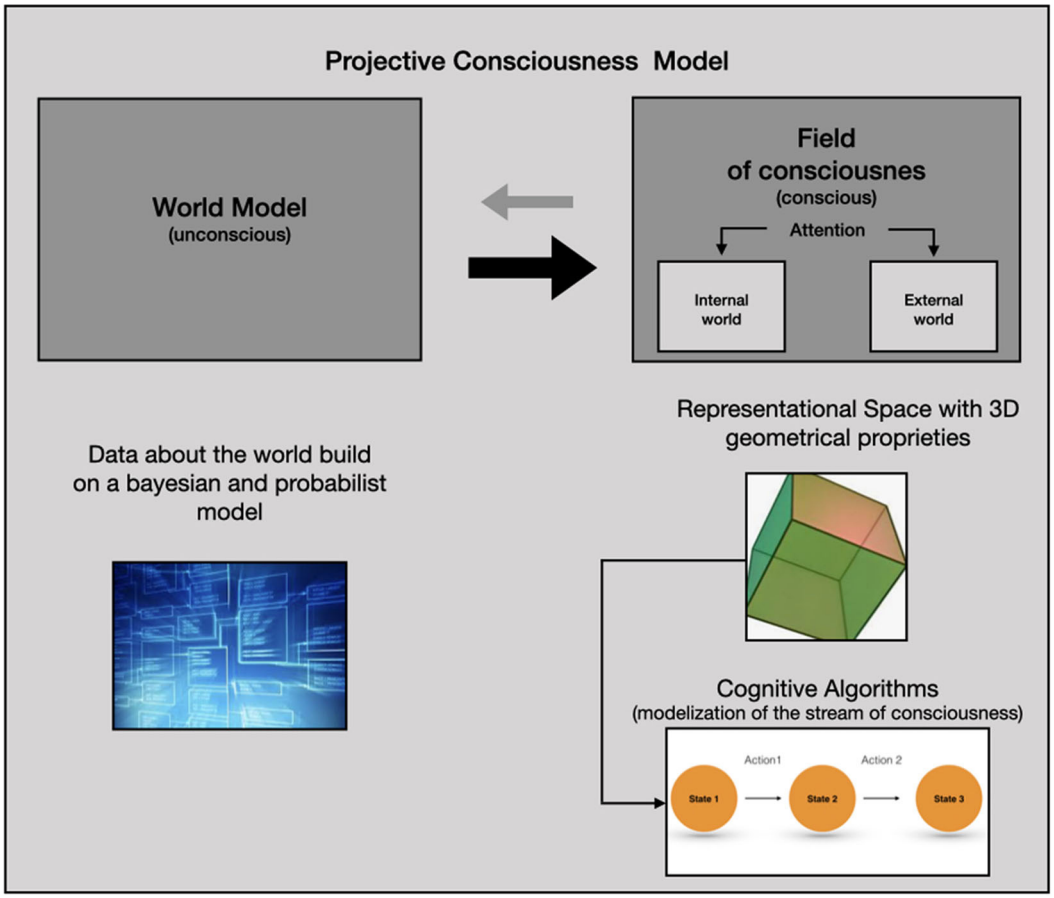

FIGURE 1 | The Projective Consciousness Model (PCM) is made of two components: (1) the world model is a set of data "stored" in the brain concerning the world which is constantly updated according to the subject's interactions with the environment (gray arrow from right to left). (2) the PCM compute a field of consciousness from the World Model on which the subjective experience is produced as a three-dimensional geometric space (black arrow from left to right, which is bolder since the FoC is computed using the World Model data). The content of this field depends on the attentional processes that can be oriented on the internal or the external world. Cognitive analysis describes the cognitive algorithms that characterize the subjective experience in the FoC.

because they have the ability to anticipate their environment thanks to these complex and reliable algorithms. Consequently, the gap between their internal representations and the actual states of the world is very limited and the resulting free energy induced by the environment decreases (i.e., a pilot is able during an accident to use specific cognitive algorithms composed of mental representations and physical behavior that he will apply in an efficient manner thanks to his training). Given that "a key function of a mind/brain is to process information so as to assist the organism that surrounds it in surviving, and that a successful mind/brain will do so as efficiently as possible" (Wiggins, 2018, p. 13), individuals will thus naturally tend to improve the quality and the complexity of their cognitive algorithms in order to increase their adaptive abilities. We propose that a research program based on the analysis and modeling of these algorithms could lead to promising empirical discoveries in these four directions:

1/The notion of "borrowed brain" has been proposed to describe how the infant internalizes the Bayesian processes of attachment figures (Holmes and Nolte, 2019). Similarly, cognitive algorithms are probably internalized during infancy from the attachment figure's own cognitive algorithms. It could be relevant to study the different cognitive algorithms used, and probably shared, by the same members of a family and especially between the children (at different ages) and their parents.

2/Propose a "genealogy" of the development of these algorithms, which take rudimentary forms during infancy—initially focused on emotional and body experiences—to very complex versions in adulthood relying notably on words. These algorithms are probably developed according to a process of increasing complexity and metaphorization as it has been shown for language (Lakoff and Johnson, 2003; Lakoff, 2014).

3/Develop a psychological test to determine precisely which algorithms are usually used by an individual. This approach can be developed, for example, to study common patterns appearing in decision-making (Tellier and Finkel, 1995). We could also "extract," in a novel manner, the cognitive algorithms used by experts in a given domain in order to better transmit them

during training programs as it has already been carried out with explicitation interviews (Maurel, 2009). In this regard, clinical applications have also been developed recently in neuropsychology, clinical psychology and psychiatry relying on neurophenomenological explorations of subjective experiences (Petitmengin, 2006).

4/Evaluate the relevance of these cognitive algorithms in terms of free energy regulation as an extension of the work developed by Rudrauf and Debban (2018). These algorithms might be a "missing link" concerning the understanding of how PCM reduces free energy. We also join the hypothesis developed in the IDyoT model (Wiggins and Forth, 2015) which relies on the "the key idea that the biggest reduction in entropy corresponds with the maximum information gain, and so the most efficient decision tree is the one that repeatedly makes the biggest possible information gain first" (Wiggins, 2018, p. 14). From this point of view, creativity could be conceived has the ability to produce original cognitive algorithms whose main function would be information efficiency and thus the reduction of free energy. Compared to the IDyoT model, CA concerns a meta-level of information treatment because it analyzes the components of the subjective experience and not the way these components have been initially produced by the brain.

These four research perspectives, relying on recent development of the Bayesian brain models and Cognitive Analysis, might open innovative perspectives both in terms of research and clinical applications. It could also help to diminish the current gap (Lutz and Thompson, 2003) in our knowledge between the first-person and the third-person point of views concerning our understanding of consciousness and subjectivity.

AUTHOR CONTRIBUTIONS

TR wrote the first draft of this paper and AF improved this draft. AF has created the CA and has described its principles. TR proposed the idea that cognitive algorithms could reduce free energy. All authors contributed to the article and approved the submitted version.

REFERENCES

Baars, B. J. (2005). Global workspace theory of consciousness : toward a cognitive neuroscience of human experience. *Progress Brain Res.* 150, 45–53. doi: 10.1016/S0079-6123(05)50004-9

Baars, B. J., and Franklin, S. (2009). Consciousness is computational : the LIDA model of global workspace theory. *Intern. J. Mach. Consciousness* 1, 23–32. doi: 10.1142/S1793843009000050

Bitbol, M., and Petitmengin, C. (2017). "Neurophenomenology and the micro-phenomenological interview," in *The Blackwell Companion to Consciousness*, 2nd Edn, eds M. Velmans and S. Schneider (Chichester: Wiley & Sons), 726–739. doi: 10.1002/97811191323

Carhart-Harris, R. L., and Friston, K. J. (2010). The default-mode, ego-functions and free-energy: A neurobiological account of Freudian ideas. *Brain* 133, 1265–1283. doi: 10.1093/brain/awq010

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. St. Lucia: Oxford University Press.

Dehaene, S. (2014). *Le Code de la Conscience*. Paris: Odile Jacob.

Dennett, D. C. (1991). *Consciousness Explained*. New York, NY: Little, Brown and Co.

Ehresmann, A. C., and Vanbremeersch, J. P. (2009). MENS a mathematical model for cognitive systems. *J. Mind Theory* 2, 129–180. doi: 10.3390/e14091703

Finkel, A. (1992). Une formalisation de l'expérience subjective. *Rapport* 92-4 du LIFAC.

Finkel, A. (2017). L'analyse cognitive, la psychologie numérique et la formation des enseignants à l'université. *Prat. Psychol.* 23, 303–323. doi: 10.1016/j.prps.2017.05.006

Finkel, A., and Tellier, I. (1996). A polynomial algorithm for the membership problem with categorial grammars. *Theoret. Comp. Sci.* 164, 207–221. doi: 10.1016/0304-3975(95)00211-1

Fodor, J. A. (1975). *The Language of Thought*. Boston, MA: Harvard University Press.

Fodor, J. A. (1979). *Representations: Essays on the Foundations of Cognitive Science*. Cambridge: MIT Press.

Fodor, J. A. (1983). *The Modularity of Mind*. Boston, MA: MIT press. doi: 10.7551/mitpress/4737.001.0001

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301. doi: 10.1016/j.tics.2009.04.005

Hobson, J. A., Hong, C. C.-H., and Friston, K. J. (2014). Virtual reality and consciousness inference in dreaming. *Front. Psychol.* 5:1133. doi: 10.3389/fpsyg.2014.01133

Holmes, J., and Nolte, T. (2019). "Surprise" and the bayesian brain: implications for psychotherapy theory and practice. *Front. Psychol.* 10:592. doi: 10.3389/fpsyg.2019.00592

Husserl, E. (1913). *Idées Directrices Pour une Phénoménologie pure et Une Philosophie Phénoménologique (1950e éd.)*. Paris: Gallimard.

Kosslyn, S. M., and Koenig, O. (1995). *Wet Mind, the New Cognitive Neuroscience*. Montreal, QC: The Free Press.

Lakoff, G. (2014). Mapping the brain's metaphor circuitry: metaphorical thought in everyday reason. *Front. Hum. Neurosci.* 8:958. doi: 10.3389/fnhum.2014.00958

Lakoff, and Johnson, M. (2003). *Metaphors we Live by*. Chicago, IL: University of Chicago. doi: 10.7208/chicago/9780226470993.001.0001

Lutz, A., and Thompson, E. (2003). Neurophenomenology integrating subjective experience and brain dynamics in the neuroscience of consciousness. *J. Consciousn. Stud.* 10, 31–52.

Maurel, M. (2009). The explicitation interview: examples and applications. *J. Consciousn. Stud.* 16, 58–89.

Pearson, J., and Kosslyn, S. M. (2013). Mental imagery. *Front. Percept. Sci.* 198:9. doi: 10.3389/978-2-88919-149-9

Petitmengin, C. (2006). Describing one's subjective experience in the second person: an interview method for the science of consciousness. *Phenomenol. Cogn. Sci.* 5, 229–269. doi: 10.1007/s11097-006-9022-2

Petitmengin, C. (2009). The validity of first-person descriptions as authenticity and coherence. *J. Conscious. Stud.* 16, 252–284.

Petitmengin, C., and Bitbol, M. (2009). Listening from within. *J. Consciousness Stud.* 16, 363–404.

Pin, J.-E. (2006). "Algorithmique et programmation. Automates finis," in *Encyclopédie de L'informatique et des Systèmes D'information* (Vuibert).

Rabeyron, T. (2020). *Clinique des Expériences Exceptionnelles*. Paris: Dunod.

Rabeyron, T., and Caussie, S. (2016). Clinical aspects of out-of-body experiences: trauma, reflexivity and symbolisation. *L'Évol. Psychiatr.* 81, e53–e71. doi: 10.1016/j.evopsy.2016.09.002

Rudrauf, D., and Debban,é, M. (2018). Building a cybernetic model of psychopathology: Beyond the metaphor. *Psychol. Inquiry* 29, 156–164. doi: 10.1080/1047840X.2018.1513685

Solms, M., and Friston, K. (2018). How and why consciousness arises: some considerations from Physics and physiology. *J. Consciousn. Stud.* 25, 202–238.

Solms, M. A. (2013). The conscious id. *Neuropsychoanalysis* 15, 5–19. doi: 10.1080/15294145.2013.10773711

Tellier, I., and Finkel, A. (1995). "Individual regularities and cognitive automata," in *Fourth International Colloquium on Cognitive Science* (ICCS 95). Donostia-San Sebastian.

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Paris: MIT press. doi: 10.7551/mitpress/6730.001.0001

Vermersch, P. (2012). *Explicitation et Phénoménologie: Vers une Psychophénoménologie*. Paris: PUF.

Wiggins, G. A. (2018). Creativity, information, and consciousness: the information dynamics of thinking. *Phys. Life Rev.* doi: 10.1016/j.plrev.2018.05.001. [Epub ahead of print].

Wiggins, G. A., and Forth, J. (2015). "IDyOT: a computational theory of creativity as everyday reasoning from learned information," in *Computational Creativity Research: Towards Creative Machines*, eds T. R. Besold, M. Schorlemmer, and A. Smaill (Amsterdam: Atlantis Press), 127–148. doi: 10.2991/978-94-6239-085-0_7

Williford, K., Bennequin, D., Friston, K., and Rudrauf, D. (2018). The projective consciousness model and phenomenal selfhood. *Front. Psychol.* 9:2571. doi: 10.3389/fpsyg.2018.02571

Wolper, P. (2006). *Introduction à la Calculabilita*. Paris: Dunod.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.