CONSCIOUSNESS MODELED: REIFICATION AND PROMISING PLURALISM

RASMUS GRØNFELDT WINTHER
University of California (Santa Cruz)

1. Introduction

A tremendous opportunity lies in store for those of us interested in consciousness. I here bracket the arguments of those who wish to minimize or reduce consciousness away, either up into socially-based discourses and institutional power (e.g., Luhmann or Foucault) or down into sheer neural patterns (e.g., Paul Churchland). Rather, consciousness will be understood as a phenomenon and process requiring study. What is paradoxical is that explorers of the territory of consciousness seem to be studying consciousness out of existence, from inside the field of «consciousness studies». How? Through their love of the phenomenon/process, they have developed powerful single models – lenses – through which to understand consciousness. But in doing so, they also seek to destroy the other *equally useful* lenses. Our opportunity lies in halting the vendettas and cross-speakings/cross-fire. The imploration is to stop the dichotomous thinking and pernicious reification of single models, and instead search for divisions of labor, complementarities, and legitimate redescriptions among the various extant models.

In other words, what would happen if we reimagined the conceptual classifications of the various models of consciousness, classifications based on general philosophical dichotomies (e.g., representational/non-representational and individualist/non-individualist), as a variety of compatible and even complementary perspectives on the same complex phenomenon and process? What would happen if rather than dig in our heels vis-à-vis our favorite theory of consciousness, at the exclusion of all the others, we saw our perceived enemy as an actual, indeed necessary, friend-in-waiting? What would it take to see a battlefield as a collaborative opportunity, to see a promising pluralism rather than an endless state space of conflict? This paper is a brief exploration of, almost a prolegomenon to, these questions.

In the next section, I present compact descriptions of three models of consciousness: *computational*, *networked*, and *embodied*. The last section fleshes out *promising pluralism*. I argue that multiple models are needed to explain, understand, and intervene in complex phenomena:

1. They *divide (and complement) ontological labor* in that multiple models focus on different parts and aspects of consciousness.
2. They *divide (and complement) explanatory labor* by asking different questions, employing distinct ontologies, and using distinct methods to understand consciousness and its parts and aspects.

3. They sometimes provide *legitimate redescriptions* of each other, which itself increases understanding.

Let us now turn to our descriptive task.

2. Three Models of Consciousness: Computational, Networked, and Embodied

The three models to be investigated are each built on a «central metaphor» [1]: *computers*, *networks*, and *the body*. Each of these three *things*, both abstract and concrete, provide a wealth of associations, assumptions, and dichotomies (both «productive» and «pernicious»), and an object/process of study for each model of consciousness. It is my joy here as a student of these models to engage in what I call an «assumption archeology» (following Michel Foucault, *Les mots et les choses*, Michael Friedman, 1999, and Ian Hacking, 2002; see also Winther, 2012), which is a sort of conceptual excavation of the hidden presuppositions and constitutive principles that stand behind or under or within a given model. The second and third models of consciousness are related, but *they emphasize different missing aspects of computational consciousness: (potentially representational, but anti-individualistic) extended-ness and (potentially individualistic, but anti-representational) materiality, respectively*.

2.1. *Computational Consciousness*

According to the first computational model of consciousness (CC, hereafter), reasoning, thinking, and reflection consist of three components: (1) representations that are (2) formal and (3) manipulated/transformed according to explicit rules (themselves internal to the representations). The mind as a computer (i.e., a Turing Machine, a Finite State Machine) is the dominant metaphor under CC. Let us see how by discussing each component in turn.

What follows is a description of consciousness and the world according to CC. *Representations* (or models) of the world are both (i) abstracted from sense-data and (ii) programmed with in-built categories and rules (e.g., Scott Atran's cognitive modules of natural kinds, or Chomsky's universal grammar – each module or the grammar(s) is full of internal assumptions and biases about the world). The world according to CC, is a complex interplay between empirical regularities, including causal relations and robust processes and stable objects, *and* problems to be solved in the world. It includes the names of your grandmothers and memories of them, as well as concerns with where you will stay and what you will see when you travel to South Africa. Sec ond, representations are *formal*. That is, they are expressed in strings of symbols. Third, representations as strings of symbols (e.g., program

[1] See Galison (1988) who develops an intercalated brick-wall «central metaphor» with three layers, data, instruments, and theories.

algorithms in C++ or sets of mathematical functions expressed in differential calculus, which could themselves be translated into algorithmic programs) are themselves subject to precise and highly-constrained *transformation rules*. But where are these rules represented? They are themselves programmed in the rich representations. Some symbol strings contain information about various aspects of the world. They are «world models». Such symbols have explicit meanings – no latent, tacit or hidden meanings or knowledge here. Others are the rules for manipulating the world models (i.e., the programs). Put differently, in terms of cognitive content and manipulation, there is nothing outside the formal representations.

Let us turn from (brief) analysis to (brief) history. It is well known that Behaviorism was the leading school of thought in psychology in America from the 1930s to 1960s. This tradition had a strong effect in academia across the globe (although Psychoanalysis and Gestalt Psychology, among other alternatives, also survived and continued developing in Continental Europe, among other places). The internalist break with Behaviorism came in the late 1950s. *Computers* were the source of the new cogntivist metaphor of mind and consciousness. The development of computers, and the existence proof of programs demolished the cognitive allergies of the behaviorists. Back then, those who believed in internal cognitive dynamics were the underdog and their new views on the strong analogies between human abstraction, reasoning, and problem-solving, on the one hand, and computer program calculations, compiling, and problem-solving inspired a significant amount of new work and young minds.

Three select quotes and two book titles will motivate this high-level history. First the quotes. In a short, elegant piece started in 1954, the Artificial Intelligence pioneer Marvin Minsky, wrote: «mental processes resemble more the kinds of processes found in computer programs: arbitrary symbol-associations, treelike storage schemes, conditional transfers, and the like» («Matter, Mind, and Models», 1965). In his influential book, *Gödel, Escher, Bach*, Hofstadter characterized consciousness thus: «Consciousness is that property of a system that arises whenever there exist symbols in the system which obey triggering patterns somewhat like the ones described in the past several sections [p. 385; the «triggering patterns» referred to are all low-level, structured neuronal activation networks]» [2]. Third, a recent comment on the cognitivist revolution is instructive: «The cognitive revolution overcame concerns about inner representational states in large part by pointing to computers. Computers operate autonomously on the basis of inner representation states, without falling into an explanatory regress, so why not people?» (Clapin, 2002; introduction, p. 13). CC broke with behaviorism by acknowledging (or reifying?) the existence of formal representations which it simultaneously placed *inside* the mind and *above* materiality.

[2] GEB (as it is also known) inspired many, and became the Book for a number of undergraduates majoring in programs such as Symbolic Systems, started in the 1980s at Stanford University. As a student at Stanford in the 1990s, I directly witnessed how powerful CC was.

Now recall two book titles. A well-known textbook in computer science, *Algorithms + Data Structures = Programs* (Wirth, 1975), says it all in the title. Programs (CC: representations) consist of algorithms (CC: transformation rules) and data structures (CC: abstracted empirical regularities). Second, Chomsky's famous transformational grammar rebuttal of behaviorism, *Syntactic Structures* (1957), tied representational internalism and formalism in the title itself. The computer scientist Knuth, so important to the analysis of algorithms and the creator of the typesetting program *TeX*, had this to say of Chomsky's book: «Here was a marvelous thing: a mathematical theory of language in which I could use a computer programmer's intuition!» (2003, preface) (Interestingly, Knuth read *Syntactic Structures* on his honeymoon.) In short, under CC, consciousness consists of representations that are programs, constituted by algorithms (rules) and data structures [3].

All revolutions start with hope. The cognitivist revolution is no exception. But has CC ended in despair or hype? It still has extremely influential proponents (e.g., Jerry Fodor, Zenon Pylyshyn, and developers of the ongoing AI project, *Cyc*). But CC as a model of consciousness has weakened for two related reasons: internal difficulties and the increased availability of alternatives. Regarding the former, issues such as the «framing problem», which inquires into how a program can pick out only the *relevant* inputs and outputs from the myriad possibilities of sense and rule inputs, and behavioral and/or problem-solving outputs, or the exceedingly great difficulty programs have with modeling and effecting bodily movement, have made CC a rather unpopular candidate for explaining consciousness. In addition, alternative models, to which we shall now turn, have themselves ushered in their own revolutions. My plea, however, is for attempting to locate places where these models might mutually strengthen one another.

2.2. *Networked Consciousness*

A classic exercise in analytical epistemology asks whether the instruments we use to gather sense-data (e.g., telescopes, microscopes, Geiger counters) could count as sensory apparatus if we hypothesized complex organic beings with analogous internally-built detecting capabilities [4]. That is, if extraterrestrial beings were directly able to see pulsars pulsating in the deep recesses of outer space, or bacterial cells in pond water, would that sense data be «internal» or «external» to the beings? Some epistemologists (or undergraduates who have just been asked the question), argue that in this hypothetical thought experiment, the sense data is indeed internal to the beings, whereas for us humans it would remain external. Others argue that this thought experiment shows that there is

[3] A related, complementary analysis of CC is found in the Stanford Encyclopedia of Philosophy entry on «The Computational Theory of Mind» (CTM) (Horst, 2011). Most basically the article argues that CTM = Representational Theory of Mind + Computational Account of Reasoning. Again, Consciousness (Mind) is fleshed out in terms of the formal representations and their manipulation.

[4] Thanks to Judith Baker (York University) for reminding me of this thought-experiment.

no hard and fast line between internal and outer, and hence much possible sense data is either wholly internal or external (depending on whether you are an internalist or externalist). A third group claims that gedankenexperiments are effectively meaningless and useless as intuitions always differ, and thus the natural sensory apparatus of those who practice science (i.e., *Homo sapiens*) is what determines the boundary between internal and external. These are clearly nuanced epistemological and semantic disputes.

Exactly analogous disputes are carried out in the arguments over consciousness. The Cartesian homunculus is a single, individualistic creature, somewhere inside each and every one of our (literal) heads. Typically, CC holds it to do reasoning and thinking by itself, without assistance from the outside world. But could parts of the external world, outside of the individual's head (e.g., Otto's notebook, to which we will turn below), somehow be part and parcel of consciousness itself? Networked Consciousness (NC) says *yes*, and is thus an externalist member of the second group referred to above. In contrast to the rationalistic and individualistic Cartesian and Enlightenment view, NC insists that «the homunculus» is neither alone nor single. S/he is not well-individuated. Rather, consciousness is extended. Abstraction and computation (or whatever your favorite version of thinking is, representational or anti-representational) is a *distributed process*. NC is not necessarily anti-representationalist, but it is anti-individualist. Let us explore two important contributions to NC, Clark and Chalmers' concept of «the extended mind» (1998) and Hutchins' (1995) notion of «distributed cognition».

In their ground-breaking essay «The Extended Mind», Clark and Chalmers invited us to think of cognition and even consciousness as a distributed process. They still thought that there was something special about what happened inside the head (which we cannot fully elucidate in this article), but their main point was to extend mental and cognitive processes to outside the skull. Their central example is Otto and Inga, two figures who need to find their way to MoMA in New York City. They are both far away, in other parts of NYC. Otto suffers from Alzheimer's disease and must carry around a notebook in which he writes many pieces of information. Among these, are the directions to MoMA from pretty much any point in NYC. In contrast, Inga is perfectly healthy and relies on her «internal memory» to get to MoMA from any point in NYC. Clark and Chalmers ask whether the notebook is in any sense part of Otto's own memory, reasoning, and cognition. They write: «In both cases [Otto and Inga] the information is reliably there when needed, available to consciousness and available to guide action, in just the way that we expect a belief to be.» There is no in-principle difference between the two sources of information (Otto's notebook, Inga's head). Indeed, «there is nothing sacred about skull and skin.» They summarize their extended mind perspective thus:

> «… the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. All the components in the system play an active causal role, and they jointly govern behavior in the same sort of way that cognition usually does. If we remove the external component the system's behavioral competence

> will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head».

Clark and Chalmers' paper was extremely influential and brought home the point that many aspects of consciousness and cognition should be understood as distributed and *networked*.

Slightly earlier work on this topic had made a similar point by using a sort of philosophical anthropology approach. In *Cognition in the Wild*, Hutchins had explored how information flowed and decisions were made in airline cockpits. In a co-authored paper (Hollan *et al.,* 2000), Hutchins summarized his view thus:

> «Whereas traditional views look for cognitive events in the manipulation of symbols inside individual actors, distributed cognition looks for a broader class of cognitive events and does not expect all such events to be encompassed by the skin or skull of an individual. For example, an examination of memory processes in an airline cockpit shows that memory involves a rich interaction between internal processes, the manipulation of objects, and the traffic in representations among the pilots. A complete theory of individual memory by itself is insufficient to understand how this memory system works. Furthermore, the physical environment of thinking provides more than simply additional memory available to the same processes that operate on internal memories. The material world also provides opportunities to reorganize the distributed cognitive system to make use of a different set of internal and external processes.» (pp. 175-176).

Hutchins and collaborators emphasized the distributed networks aspects of cognition and consciousness. They also explored the different modalities and mechanisms of cognitive processes. That is, by distinguishing among, for instance, «internal processes», «manipulation of objects», and «traffic in representations», these scholars moved towards a taxonomy of different types of processes and objects involved in extended or distributed cognition. Not all abstraction and thinking is the same, nor are all parts of cognitive processes identical or even similar in type. The different roles actual computation plays, as compared to information stored in physical devices such as altimeters or pressure gauges, as compared to conversation and information exchange across the pilot, co-pilot, flight engineer and so forth needs to be analyzed. An important contribution of Hutchins' work is to start working towards a «typology of components of distributed consciousness.» It is not sufficient to say that consciousness is distributed, we must also explore the structure and function of its parts.

NC takes issue with the individualism of CC. It extends consciousness to include what is outside the individual. That is, the scaffolding of consciousness is itself part of consciousness. This anti-individualist move problematizes the internal vs. external dichotomy, and perhaps even overcomes it [5]. Moreover, NC organizes the different components of this extended system into a taxonomy.

[5] See Oyama (2000) for an analogous move to overcome the internal/external dichotomy in the nature/nurture debates so entrenched in the biological and psychological sciences.

Distinct parts have specific structures and functions. The research program of the NC model of consciousness thus involves characterizing *parts* and *roles* of *an extended consciousness system*. In short, we must explore extended relationality in our efforts to understand consciousness.

2.3. *Embodied Consciousness*

A different set of traditions, both philosophical and empirico-scientific (and their combination), have also resisted the CC model of consciousness. The Embodied Consciousness (EC) alternative model of consciousness is defended in a broad (and sometimes conflicting) variety of ways by investigations as different as (1) the diverse phenomenological tradition of Edmund Husserl, Martin Heidegger, Maurice Merleau-Ponty, and Alfred Schutz, (2) the more unified pragmatism of William James and John Dewey, and (3) the recent empirico-scientific studies of embodied robotics by Rodney Brooks and of embodied linguistics by Mark Johnson [6]. One common thread across these traditions is a focus on the body, and on imagining thinking, cognition, and reasoning as embodied rather than abstract processes. In what follows, I will briefly motivate the general problem of CC, as seen from the EC model. I will then turn to three *modalities* or *parts* of consciousness highlighted by EC and strongly downplayed by CC: (1) (sophisticated) sense data, (2) movement, and (3) feelings. In order to itself embody these modalities/parts of consciousness, we will meet cutting-edge work in robotics, linguistics and kinesthetics [7].

Philosophically, the incarnation of CC is the Cartesian homunculus. It is well-known that Descartes posited an ego that cogitated. Indeed, the only principle of which Descartes could be certain in the bout of systematic doubt with which he starts his *Meditations* (1641), was that there was an *I* doing all the thinking, all the doubting. This thinking *I* consisted of *res cogitans* – i.e., cogitating thing/stuff/matter. Upon this *I*, Descartes built his system of knowledge. Some have since imagined this I as a sort of reified homunculus, an abstract representer of abstractions, deeply severed from his (almost invariably and importantly a «him») body, and processing raw sense-input with his abstract central processing unit. Almost needless to say, Kant built further on Descartes picture of an abstract, individuated, severed reasoning agent by postulating a rich, inner, structure rationality. Kant's philosophy gave us an understanding of the 'homunculus' (if we may) internal conceptual/rational categories, unity of apperception, and bounds of reason.

---

[6] Two other essential figures, with well-articulated research programs are Antonio Damasio (1994) and Maxine Sheets-Johnston (1999). The former explores a neuroscience focusing on embodiment and feeling; the latter investigates the literal role of movement and kinesthetics in self-awareness. Unfortunately, there is not sufficient space-time to investigate these important views (see also MacIver, 2009).

[7] Wilson (1995) and Anderson (2003) provide ample discussion regarding the NC and EC models of consciousness.

This homunculus has been resisted. One of the ways its existence was questioned, and its de-reification invited, was by emphasizing the corporeal parts and modalities of consciousness. Mind was not separate from body – indeed it was shaped by, constrained by, and guided by the whispers of physicality. In his monumental *Being and Time* (1927), Heidegger wrote:

> «The kind of dealing which is closest to us is as we have shown, not a bare perceptual cognition, but rather that kind of concern which manipulates things and puts them to use; and this has its own kind of 'knowledge'» (Heidegger, 1962, p. 95).

As has so been so ably analyzed by the critic of AI (Artificial Intelligence) and CC, Hubert Dreyfus (e.g., 1992), Heidegger is an excellent point source for reflections about human being-in-the-world. Heidegger's work which downplay representation and abstract thinking, and instead focuses on our material activity, on our use of tools-for particular purposes, and on our literal bodies.

Related resistance to Descartes' homunculus is found in William James' evolutionarily-aware and anti-«vicious abstractionist» (James' own term from *The Meaning of Truth* 1909/1911) meditations on psychology:

> «Mental facts cannot be properly studied apart from the physical environment of which they take cognizance. The great fault of the older rational psychology was to set up the soul as an absolute spiritual being with certain faculties of its own by which the several activities of remembering, imagining, reasoning, willing, etc. were explained, almost without reference to the peculiarities of the world with which these activities deal. But the richer insight of modern days perceives that our inner faculties are *adapted* in advance to the features of the world in which we dwell, adapted, I mean, so as to secure our safety and prosperity in its midst» (James, 1900, p. 13).

James had published his justly famous *Principles of Psychology* in 1890. In emphasizing (his term) «the stream of consciousness» and the inseparability of sentiment and rationality, James lay the groundwork for an empirically-grounded *embodied* model of consciousness, an alternative to the CC model of consciousness. Evolutionary theory entered in two ways. First, by emphasizing our continuity with animals, including continuity in sensorimotor apparatus and processes, and continuity in problems and issues-going-of-concern relevant to survival. Second, by highlighting the continuity, indeed interpenetration, of fact and value. The way we are, and behave in the world, are inseparably tied to the way we *should* be and behave. James did find ways to use evolutionary theory, and our adapted and material bodies, as a normative platform from which moral rules emanated. These rules neither diffused into unjustified relativism nor ware they Kantian-like idealistic and rational top-down moral imperatives [8]. In short, both Heidegger and James are key figures from

[8] Useful essays on James can be found in Putnam 1997. I am fortunate to be a member of Lucas McGranahan's PhD committee (Department of Philosophy, University of California [Santa Cruz]). Lucas is comparing Nietzsche and James, with a particular eye on their respective

Phenomenology and Pragmatism, respectively, that resist the Cartesian homunculus by adopting the EC model of consciousness.

Scientific work which could be said to fall within embodiment theory includes the robotics of Rodney Brooks and the linguistics of Johnson[9]. Each of these research programs studies related sets of modalities or parts of consciousness, not studied by CC (nor by NC, a non-individualist model). Brooks is interested in (sophisticated) sense data and movement, which are part of a strong robot-environment coupling. Johnson is particularly concerned with human movement and feelings, as captured in, and determining of, our dominant metaphors. Let us briefly explore each.

As discussed in his intellectual biography (2002), Brooks has worked on robotics for a number of years, using a paradigm utterly different than the representationalist one typical (ay, definitional) of CC. He decided to forego representations altogether, and instead build tight robot-environment information loops, in which the ant-like robot reacted immediately to feedback from its environment and adjusted its behavior accordingly. In *Cambrian Intelligence*, Brooks writes: «Essentially the idea is to set up appropriate, well conditioned, tight feedback loops between sensing and action, with the external world as the medium for the loop» (p. 109). In his now-classic paper «Intelligence without representation», he wrote:

«We have reached an unexpected conclusion (C) and have a rather radical hypothesis (H):

(C) When we examine very simple level intelligence we find that explicit representations and models of the world simply get in the way. It turns out to be better to use the world as its own model.

(H) Representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems.»

In other words, the world is its own best model and representation is an inappropriate abstraction unit for building an (artificially) intelligent system – «in order to understand something, I must build it», might be an engineer's motto. The «body» of the robot uses complex sensory data and highly reactive motor units to deal effectively with its environment. Tacit knowledge (Michael Polanyi's concept) itself gets entrenched and built in a bottom-up fashion from robot-environmental interaction. Brooks' robots are a beautiful exemplar, in the kuhnian sense, of the EC model. Analogously, our own bodies are highly sensitive

views on evolution and ethics. His work on James has helped me see the founding figure of psychology in the USA with a fresh perspective (see McGranahan, 2011).

[9] I also work with Alexis Mourenza, another PhD student at the University of California (Santa Cruz). Her detailed conceptual and scientific investigations of animal consciousness point to a whole other area of research closely tied to the EC model which I must also side-step in this paper due to space limitations. See, for example, Bekoff, Allen, and Burhgardt (2002).

reactive units to our environment. Our consciousness is grounded in and shaped by the embodied flow of information, sensations, and qualia.

Mark Johnson is a linguist and philosopher perhaps best known for his co-authored book *Metaphors We Live By* (1980, Lakoff and Johnson) with George Lakoff. In a more recent book, *The Meaning of the Body*, Johnson explores EC, with an eye towards the process of aesthetic judgment and establishment of meaning in our lives. In one particularly constructive chapter, «From Embodied Meaning to Abstract Thought», Johnson fleshes out how abstract thought, reasoning, and consciousness could arise from the various parts, processes, and movements of the body. Drawing on his linguistic studies, Johnson argues that our materiality is often the source for our abstract concepts and thinkings. There is a «conceptual metaphor» (his term) structure built into the relations between our bodies and our languages. In this chapter, he writes:

> «Dewey's pragmatist continuity thesis claims that we must be able to move, without any ontological or epistemological rupture, from the body-based meaning of spatial and perceptual experience that is characterizable by image schemas and affect contours all the way up to abstract conceptualization and reasoning. This same notion of ontological continuity underlies most second-generation (embodied) cognitive science. The existence of *abstract* concepts thus poses a fundamental problem for any naturalistic view of meaning as grounded in the qualities and structures of sensorimotor experience. How can thinking about abstract, nonphysical entities possibly be grounded in the body?» (p. 176).
>
> …
>
> «… abstract concepts are defined by conceptual metaphors that recruit the semantics and inference patterns of sensorimotor experience. … AFFECTION IS WARMTH, IMPORTANT IS BIG, MORE IS UP/LESS IS DOWN, HAPPY IS UP/SAD IS DOWN, STATES ARE LOCATIONS, CAUSES ARE FORCES, … TIME IS MOTION…» (pp. 178-179).

Conceptual metaphorical structures take body experiences, feelings and movements as the *source* [e.g., warmth, spatial orientation, and (literal) forces], and abstract thoughts and reasonings are the *target* (e.g., affection, more or less, and causes/causation). Johnson holds that we could generalize conceptual metaphors to many other aspects of our abstract toolbox. He also appeals to other sorts of embodied reasoning processes, where the body constrains and shapes our thought.

EC is a different, but related (to NC) way, of reacting to CC. Here the focus is on internal, material modes of having and experiencing and making consciousness. In particular we look at the body, and its (1) (sophisticated) sense data, (2) movement, and (3) feelings. By submerging into the body, we explore the relationship between body and environment, and see that body and consciousness are deeply enmeshed and intertwined. CC's view from above is hardly sufficient.

In this second section, I have provided brief descriptions of the models of consciousness under our purview here. An *assumption archeology* has been carried out on each of them. Needless to say, much work remains to be done.

3. Promising Pluralism

The goal of our exercise has been to emphasize three models of consciousness, with an eye towards their respective division and complementation of labor, and their ultimate integration. Consciousness is a complex process, with at least three loci, as described here: (1) representations that can be described formally, (2) externalized, networked components, (3) embodied situated-ness. That is, consciousness has to be understood through at least a trichotomy which we must overcome: formal representation(above) vs. embodied(below/inside) vs. networked(outside). This turns out actually to be a trichotomy (formal representation/embodied/networked) correlated with various dichotomies both general (e.g., representational/non-representational and individualist/non-individualist) and more specific (e.g., representational/embodied and individualist/networked). The standard position on dichotomies (or trichotomies or *n*-chotomies), is to see the opposing poles or ends as mutually exclusive and collectively exhaustive. However, we can also view these standard oppositions as an opportunity for dialectical overlap and interpenetration (see, e.g., Levins and Lewontin 1985, Winther 2008, 2011). No single model should be individually and imperialistically reified. Rather, there is strength in dialogue and numbers (see also Mitchell 2009). I conclude this prolegomenon on models of consciousness by stating three lessons.

The first lesson is that each model is important and relevant. Each emphasizes distinct parts and aspects of the ontology of consciousness and of explanations pertinent to consciousness. Each asks different questions, and employs its own methodologies. CC is not sufficient. It lacks an account of feeling, sense-data, and movement – the three essential components of EC – which influence both the form and content of consciousness. Moreover, it is too internalistic and individualistic. It lacks an account of the variety of external processes and objects – elucidated by NC – which again necessarily modulate and shape consciousness. Each model of consciousness is necessary and insightful. Complex consciousness, like the proverbial elephant examined by the blind men of the Eastern fable, requires analyses from different points of view. A (static) pluralism is thus necessary to get a complete understanding of consciousness.

But there are more and deeper lessons. In our attempts to avoid and to overcome reification, which is the way I prefer to describe the purpose of this paper, we (1) engage in dialogue, (2) are sensitive to data, (3) engage in self-reflection (our «assumption archeology»), and (4) attempt integration. Through these activities, we see that especially in this case of modeling consciousness, the three models are not only individually necessary, but they also shape and constrain one another. That is, they exhibit a definitional and dynamical dialectic. For instance, representations are vulnerable to emotions and sense-data, and to inputs being received from a broad variety of external places. The dichotomies are vulnerable and contingent. The second lesson is that *dialectical thinking* is useful here. Definitions as well as actual process-dynamic explanations of each

model of consciousness are sensitive to the definitions and explanations of the other two models. Not only is each model necessary, each model constrains and shapes the other two. A (dynamic) pluralism thus itself changes its own grounds of possibility.

The third lesson is that our classification is itself unstable and vulnerable. It can change with the introduction of *further* models of consciousness. For instance, introducing Hameroff and Penrose's quantum consciousness, would add a whole new physical theory to the physical substrate of consciousness. This addition would invite us to see new triangulations, new aspects, of the models of consciousness we have thus far explored. *The «consciousness wars» are better thought of as «consciousness collaborations»*. Like any part of science, work on the sciences of the mind and consciousness is a Ship of Theseus that we are constantly rebuilding at sea. Such reconstruction requires collaboration.

Acknowledgements

Thanks to Jens Degett and Cristina Heller del Riego for inviting me to a thought-provoking scientific and philosophical conference. Thanks to Quentin Cooper, Stuart Hameroff, Amparo García-Plaza, Juani Guerra, Lluis Oviedo, and Steen Rasmussen, among many others at the «Life, Evolution, and Complexity» conference (December 2010, Comillas University, Madrid), for conversation. Given the context of the enjoyable conference at which this paper was presented, I believe it instructive to self-critically state my assumptions regarding religion. I am not religious in any standard sense, and certainly do not believe in any monotheistic deity. I am an avid supporter of scientific methods, and of careful philosophical discussion of the empirical and ethical strengths and weaknesses of the products (e.g., models, theories, interventions) of such methods. I am generally agnostic about the possible (rational and justified) relationships between science and religion. Jácome Armas, Judith Baker, Mark Detweiler, Claus Emmeche, Peter Godfrey-Smith, Ian Hacking, Hervé Kieffel, Sergio Martínez, Amir Najmi, Susan Oyama, and Frederik Stjernfelt have provided a gentle forum of discussion for many of these ideas, for a number of years.

Bibliography

Anderson, M. L. (2003), «Embodied Cognition: A field guide», *Artificial Intelligence* 149: 91-130.

Bekoff, M.; Allen, C., and Burghardt, G. M., *The Cognitive Animal. Empirical and Theoretical Perspectives on Animal Cognition,* MIT Press.

Blackmore, S. (2005), *Conversations on Consciousness,* Oxford University Press.

Brooks, R. (1991), «Intelligence without representation», *Artificial Intelligence* 47:139-159. Reprinted in *Cambrian Intelligence: The Early History of the New AI*, 1999, MIT Press.

Brooks, R. (2002), *Flesh and Machines. How Robots Will Change Us*, New York: Pantheon Books.

Chomsky, N. (1957), *Syntactic Structures*, Berlin: Mouton de Gruyter.

Clapin, H. (ed.) (2002), *Philosophy of Mental Representation*, Oxford University Press.

CLARK, A., and CHALMERS, D. J. (1998), «The Extended Mind», *Analysis* 58:7-19.
DAMASIO, A. (1994), *Descartes' Error. Emotion, Reason, and the Human Brain*, New York: GP Putnam's Sons.
DREYFUS, H. (1992), *What Computers (Still) Can't Do: A Critique of Artificial Reason*, MIT Press.
FOUCAULT, M. (1970), *The Order of Things. An Archeology of the Human Sciences*, Vintage Books.
FRIEDMAN, M. (1999), *The Dynamics of Reason*, Stanford, CA: CSLI Publications.
GALISON, P. (1988), «History, Philosophy, and the Central Metaphor», *Science in Context* 2:197-212.
HACKING, I. (2002), *Historical Ontology*, Cambridge, MA: Cambridge University Press.
HAMEROFF, S. (2010), «The "conscious pilot" – dendritic synchrony moves through the brain to mediate consciousness», *Journal of Biological Physics* 36:71-93.
HAMEROFF, S., and PENROSE, R. (Online), «Conscious Events as Orchestrated Space-Time Selections». Found at following URL: http://www.quantumconsciousness.org
HEIDEGGER, M. (1962), *Being and Time*, Harper Collins.
HOFSTADTER, D. (1979), *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books.
HOLLAN, J.; HUTCHINS, E., and KIRSH, D. (2000), «Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research», *ACM Transactions on Computer-Human Interaction* 7:174-196.
HORST, S., «The Computational Theory of Mind», *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2011/entries/computational-mind/>.
HUTCHINS, E. (1995), *Cognition in the Wild*, MIT Press.
JAMES, W. (1900), *Psychology. Briefer Course*, American Science Series, New York: Henry Holt and Co.
— (1909/1911), *The Meaning of Truth. A Sequel to 'Pragmatism',* New York: Longmans, Green, and Company.
JOHNSON, M. (2007), *The Meaning of the Body. Aesthetics of Human Understanding,* University of Chicago Press.
KNUTH, D. E. (2003), *Selected Papers on Computer Languages*, Stanford, CA: CSLI Lecture Notes, no. 139.
LAKOFF, G., and JOHNSON M. (1980), *Metaphors We Live By,* University of Chicago Press.
LEVINS, R., and LEWONTIN, R. (1985), *The Dialectical Biologist*, Harvard University Press.
MACIVER, M. A. (2009), «Neuroethology: From Morphological Computation to Planning», *The Cambridge Handbook of Situated Cognition,* Robbins, P., and Aydede, M. (eds.), Cambridge University Press, pp. 480-504.
MCGRANAHAN, L. (2011), «William James's Social Evolutionism in Focus», *The Pluralist* 6(3):80-92.
MINSKY, M. L. (1965), «Matter, Mind and Models», *Proc Intl Federation of Information Processing Congress* 1:45-49. (Also easily found online: http://groups.csail.mit.edu/medg/people/doyle/gallery/minsky/mmm.html)
MITCHELL, S. (2009), *Unsimple Truths. Science, Complexity and Policy,* University of Chicago Press.
OYAMA, S. (2000), *The Ontogeny of Information. Developmental Systems and Evolution*, 2nd ed., Duke University Press.
PUTNAM, R. A. (ed.) (1997), *The Cambridge Companion to William James*, Cambridge University Press.
SHEETS-JOHNSTONE, M. (1999), *The Primacy of Movement*, Amsterdam: John Benjamins.

WILSON, R. A. (1995), *Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind*, Cambridge University Press.
WINTHER, R. G. (2008), «Systemic Darwinism», *Proceedings National Academy of Sciences*, 105 (33): 11833-11838.
— (2011), «Part-Whole Science», *Synthese* 178:397-427.
— (2012), «Evo-Devo as a Trading Zone», in: A. LOVE (ed.), *Conceptual Change in Biology: Scientific and Philosophical Perspectives on Evolution and Development,* Springer Verlag (Boston Studies in the Philosophy of Science).
WIRTH, N. (1975), *Algorithms + Data Structures = Programs,* Prentice-Hall.

RASMUS GRØNFELDT WINTHER

Philosophy Department
Karolinska Institutet, Stockholm
University of California (Santa Cruz)
Center for Philosophy of Nature and Science Studies,
University of Copenhagen
Biocomplexity Center, Niels Bohr Institute,
University of Copenhagen
rgwinther@gmail.com
www.rgwinther.com