Journal of EVOLUTION & Technology

A peer-reviewed electronic journal published by the Institute for Ethics and Emerging Technologies

ISSN 1541-0099

17(1) – March 2008

Carter's Cartesian Paraphrase and "Operational Autonomy": The Carter-Bostrom Anthropic Principle, the Principle of Mediocrity, and "Being No One. . ."

Tim Clark, Concordia University and
Hexagram: Institute for Research/Creation in Media Arts and Technologies
Journal of Evolution and Technology - Vol. 17 Issue 1 – March 2008 - pgs 59-70
http://jetpress.org/v17/clark.htm

Abstract

This paper examines Yilmaz, Ören and Aghaee's outline of present research efforts into the development of simulations that "represent the behavior of active (human) entities in the world." The paper argues that the Carter-Bostrom formulation of the anthropic principle provides a more functional set of theoretical, and pragmatic proposals to frame the issue of the simulation of human *socio*cognitive activity than the now standard conjunctive phrases "cognitive simulations," "Strong Artificial Intelligence," and "Strong Machinic Consciousness." More importantly, the principle of "anthropic entity(s)" was, in its original cosmological form, developed within the global context of evolutionary theory. Consequently, as an evolutionary principle, it may provide a more comprehensive, explanatory context with which to frame the possible emergence of the first fully operational artificial simulations. In its present formulation, however, the anthropic principle is insufficient with respect to two interlocking issues that are of foundational importance to the successful development of simulations. These are the principle of "operational autonomy," and a credible, theoretical response to two of the most generic, identitarian properties of (human) personhood: the folk psychological belief in "our" substantive, autonomous identity as persons possessing the property of being the person whom we experience as "me" or "I"; and the autocentric *preclusion* of the emergence of the concept of personhood to other forms of *emergent*, anthropic entities.

In particular, the work of Michel Foucault and Thomas Metzinger, Director of the Theoretical Philosophy Group at the Johannes Gutenberg Universität Mainz, may provide for just such theoretical augmentations of the anthropic principle.

Introduction: Intelligent agents, simulation, and gaming

To begin, some quotes:

> Recent trends have made it clear that simulation model fidelity and complexity will continue to increase dramatically in the coming decades. For example, the beginning of the mission to build a simulated brain is already announced (Graham-Rowe, 2005). (Yilmaz, Ören, Aghaee 2006: 340.)
>
> Using intelligent agents in simulation models is based on the idea that it is possible to represent the behavior of active entities in the world with their own operational autonomy. (Yilmaz, Ören, Aghaee 2006: 340.)
>
> The factors that may affect decision making of agents, such as personality, emotions, and cultural backgrounds, can also be embedded in agents. Additional abilities of agents are needed to increase their intelligence and trustworthiness. Abilities to make agents intelligent include anticipation (proactiveness), understanding, learning, and communication in natural and body language. Abilities to make agents trustworthy as well as assuring the sustainability of agent societies include being rational, responsible, and accountable. These lead to rationality, skillfulness, and morality (e.g., ethical agent, moral agent). (Yilmaz, Ören, Aghaee 2006: 344.)

The rise of the thesis of S (W)AI[1] during the twentieth century is hardly surprising, given a situation where the primary exemplar for "intelligence" was human sociocultural, linguistic, and cognitive activity. A seminal example of this process of equivocation can be found in the opening paragraph of the document, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, which was submitted to the Rockefeller Foundation in 1955, with a request for the funding to set up a conference. The document announces: "The study will proceed on the basis of the conjecture that intelligence (which is reserved for humans) can in principle be so precisely described that a machine can be made to simulate it" (McCarthy *et al.*: 1955).[2]

Some 51 years later, we find the computer scientists and software engineers, Yilmaz, Ören and Aghaee, announcing a similar objective concerning the development of "intelligent agents in simulation models." Setting aside the theoretical and pragmatic difficulties, and the immense ethico-juridical and theological issues that such a project engenders, this research raises foundational questions with regard to the nature of this "operational autonomy."

"Large Number Coincidences": Selection effects and the Principle of Mediocrity

In 1973, the theoretical physicist and cosmologist Brandon Carter presented his now famous (or *in*famous) paper, "Large Number Coincidences and the Anthropic Principle in Cosmology" (Carter 1974), at a conference to mark the five hundredth birthday of Nicholas Copernicus. The paper was ostensibly written to critique "Dirac's Large Number Hypothesis" (Dirac 1937),[3] which had been proposed by the English cosmologist Paul Dirac. In actuality, however, Carter had also directed his critique at one of the foundational principles of modern cosmology, the Generalized Copernican Cosmological Principle (Rudnicki 1995). A modern derivative of Copernicus' argument that relative to any given planetary body, the observable universe will be approximately the same, the "Generalized" version asserts that from any given point the observed universe is homogeneous and isotropic in nature. Or, conversely, according to the Principle of Mediocrity, there is no privileged point *x,* such that any observation(s) made from *x* will be privileged over any observations made from point *y.*

The later work of the Princeton theoretical physicist R. H. Dickie set in motion serious questions about

the sustainability of the Generalized Principle. In a letter to the editor of the journal *Nature*, entitled, "Dirac's Cosmology and Mach's Principle," Dickie reviewed the (Hermann) Weyl | (Arthur) Eddington | (Paul) Dirac debate over the derivation of a set of extremely large dimensionless numbers with regards to specific physical and astrophysical properties of the universe. What was important about the pure numbers for the gravitational coupling constant *G*, the Hubble age of the universe *T*, and, the mass of the universe *M* relative to its visible limits was that all had surprisingly, numerologically coincidental orders of magnitude. Why? Dickie's response was contained in a brief letter published in a 1961 issue of the journal *Nature* where he offered[4] what Carter would formulate as the argument from the "anthropic principle."

Dickie noted that if there is, as Dirac had argued, an underlying causal connection between the three numbers with respect to the physics at the quantum and cosmic levels, then for this argument to hold, one would have to hypothesize that this connection was "independent from time." The issue of the value of *T* was significant because Dirac had argued that, given the evolutionary nature of the universe, it must be expected that each of the pure numbers would vary over time. Dickie suggested that if one were to assign a random value to *T* taken from a large number of possible values, then "the present (actual) value would have a low a priori probability." Consequently, the present correspondence between the numbers "would have been highly unlikely." Dickie's proposal to resolve the issue was to argue that, given the relative evolutionary age of the universe, *T* cannot take on a large range of values insofar as they are limited by a time-constrained, reciprocal inter-relationship between the physico-biological "requirements to *make physicists*" (Dickie 1961), and the presence of these same physicists to carry out the required observations. As Abraham Zelmanov proposed in his dissertation, *Chronometric Invariants,*

> The Universe has the interior we observe, because we observe the Universe in this way. It is impossible to divorce the Universe from the observer. The observable Universe depends on the observer and the observer depends on the Universe. If the contemporary physical conditions in the Universe change then the observer is changed. And vice versa, if the observer is changed then he will observe the world in another way. So the Universe he observes will be also changed. If no observers exist then the observable Universe as well does not exist. (Zelmanov 1944.)

In 2004, at a Colloquium held at the Collège de France, Carter reflected back on his 1974 paper on "Large Number Coincidences"(Carter 2004). He used the occasion to propose a revised definition of his original formulation of "the Anthropic Principle" in light of the history of sometimes highly controversial interpretations of his proposal. In his 1974 paper, he had provided for both "Weak," and "Strong" versions of the Principle. As he stated, "these predictions do require the use of what may termed the *anthropic principle* to the effect that what we can *expect to observe must be restricted to the conditions necessary for our presence as observers*" (Carter 2004). In its "Weak" formulation, his statement asserts the trivially self-evident claim that to make any observations there has to be an observer, be that "observer" an instrument or a physicist, and of course, the presence of an observable. However, his formulation could also be interpreted as a Strong version of the principle that appeared to permit far less defensible claims such as Zelmanov's, "If no observers exists then the observable Universe as well does not exist."

In light of these issues, Carter explicitly *re*formulated it as a Bayesian "*microanthropic"* principle which asserts as follows:

> the *a priori* probability distribution for our own situation should be prescribed by an anthropic weighting, meaning that it should be uniformly distributed, not over space time (as the ubiquity principle would require), but over all observers sufficiently comparable with ourselves to be qualifiable as anthropic.[5]

The Oxford philosopher and mathematician Nick Bostrom later reformulated Carter's concept of "observer(s)": any given "observer(s) in any particular brain state (subjectively making an observation *e*)" (Bostrom 2005). Another of Bostrom's revisions was a strengthening of the Bayesian, probabilistic mode of statistical analysis with the addition of the "Self-Sampling Assumption" (SSA): "One should reason as if one were a random sample from the set of all observers in one's reference class" (Bostrom 2002). The SSA foregrounds the two central theoretical assumptions of the Anthropic Principle. The first is the use of Bayesian, probabilistic analysis, while the second is the epistemological concept of "observer(s)." The second is the most crucial component because the reference class of the concept can include any entity that functions in an observational mode, be it an instrument or a "physicist."[6] The most important member of the class is, as Bostrom notes, any "brain state (subjectively making an observation *e*)." This class is the foundational premise of the Anthropic Principle – the Argument from the First Person.

Carter's "Paraphrase" and the (Anthropic) Argument from the First Person

In the same 1974 paper, Carter, as a concluding codicil to his formulation of the Strong Principle, paraphrased Rene Descartes' famous assertion concerning reasoned thought as, "I think therefore the world is such as it is" ("cogito ergo mundus talis est") (Carter 1974: 294).

For Descartes, indefeasible epistemological claims must arise from the methodical, universal application of a hyperbolic, skeptical interrogation of the contents of, and relation between, introspective, self-reflective consciousness (*res cogitans*) and sensate experience of the body and the external world (*res extensa*). Application of hyperbolic skepticism can call into question the contents of all introspective and sensate experience; however, the very activity of the *cogito* is its own guarantee, since to deny the presence of the act of thinking would be to generate a logically contradictory claim. As Descartes notes, "So that it must, in fine, be maintained, all things being maturely and carefully considered, that this proposition (*pronunciatum*) *I am, I exist*, is necessarily true each time it is expressed by me, or conceived in my mind" (Descartes 1901).

The core of Descartes' project was initiated in the great section of the First Meditation, where he laid out the sceptical arguments concerning dreaming and non-dreaming states, and, of course, the penultimate, unbounded argument from the "Evil Genius," or the Argument from Theocentric Stability.[7] Using reasoning founded on the Thesis of Similarity with respect to the traditional Platonic argument from "epistemically distinct worlds" (Newman 2005), Descartes concluded that the Evil Genius argument was false. He argued that the sensate world, and its relation to the methodic process of thinking, were grounded in a theologically-based denial of the possibility of error and certainty. Speaking of the deity whose existence he believed he had demonstrated, Descartes wrote,

> it follows that existence is inseparable from him, and therefore that he really exists: not that this is brought about by my thought, or that it imposes any necessity on things, but on the contrary . . . the necessity of the existence of God, determines me to think in this way. (Descartes 1901.)

Consequently, at the center of First Person (Anthropic) experience is the perpetual presence of deconstructibility – of systemic doubt and error. This is the case because *if* there is no other (third person) ground than the activity of the *cogito*, then any first order knowledge claim(s) and consequent derivations drawn from these initial claim(s) must, themselves, be subject to second order constraint of the very activity of the *cogito*: the presumably *un*transcendable limiting condition on all first order experiential activity.[8] Bostrom puts this well:

> Observation selection effects are an especially subtle kind of selection effect that are introduced not by limitations in our measurement apparatuses but by the fact that all evidence is preconditioned on the existence of an observer to "have" the evidence and to build the instruments in the first place. (Bostrom 2005)

To understand the problem of error and systemic doubt is to understand that Descartes confronts two correlative epistemological issues: naive realism; and the problem of first person to third person data.[9] These two issues are structurally interrelated. The success of resolving the latter issue is derivative of our success in theoretically exposing the naivety of humanity's everyday sense of reality with regards to consciousness' interaction with the world. One might call this naivety the taken-for-granted nature of phenomenal experientiality, in that what "I" or "we" experience *is what we experience*. *Eo ipso*, two of the most important and difficult theoretical issues to emerge from Descartes' work are, can one autonomize folk psychological understandings of self-consciousness from naive realism, and two, if one is to argue successfully that one can, then that success is contingent upon the further success of the attempt to *re*ground the argument for the presence of an autonomized self-consciousness in third person data.

Famously, Descartes' realization also represented the strong – metaphysically dualist – principle of the irreducibility of the contents of mind, inasmuch as mind cannot be framed in terms of theoretical (machinic-physicalistic) explanations, or conversely, the non-extensibility thesis: that theoretical explanations cannot be extended to consciousness (*res cogitans*), inasmuch as they are restricted to the domain of nature (*res extensa*).[10] In the setting of contemporary philosophy of mind, the Australian philosopher David Chalmers has remarked, with regards to the non-extensibility thesis, that there is an "*explanatory gap* . . . between the functions and experience," with respect to the apparent dualism between the naive realism of first order, conscious experience and machinic-physicalistic explanations of human phenomenal experientiality.

This remark occurs in Chalmers' seminal 1995 paper, "Facing Up to the Problem of Consciousness," where he addresses the nature of *qualia*, the sense that when we are given explanations in terms of the physics of light, of neurobiology, of the constructionist mediation of discursive systems, we are still left with "our" experience of the seamless, immersive sense of being in the world. As Chalmers notes,

> Why is it that when electromagnetic waveforms impinge on a retina and are discriminated and categorized by a visual system, this discrimination and categorization is experienced as a sensation of vivid red? We know that conscious experience *does* arise when these functions are performed, but the very fact that it arises is the central mystery. There is an *explanatory gap* . . . between the functions and experience, and we need an explanatory bridge to cross it. A mere account of the functions stays on one side of the gap, so the materials of the bridge must be found elsewhere. (Chalmers 1995.)

"Are You Living in a Computer Simulation" and the question of autonomy

In 2003, the Swedish philosopher Nick Bostrom published his paper, "Are You Living in a Computer Simulation?" (Bostrom 2003). In his conclusion, he notes as follows:

> Based on this empirical fact, the simulation argument shows that at least one of the following propositions is true: (1) The fraction of human-level civilizations that reach a posthuman stage is very close to zero; (2) The fraction of posthuman civilizations that are interested in running ancestor-simulations is very close to zero; (3) The fraction of all people with our kind of experiences that are living in a simulation is very close to one. (Bostrom 2003.)

Bostrom's paper has generated much controversy, especially given its somewhat science-fiction-like quality. The paper itself is part of a larger debate concerning "Doomsday," or human extinction, scenarios that have their roots in Carter's work, in particular his 1983 paper, "The Anthropic Principle and its Implications for Biological Evolution." I will not rehearse Bostrom's argument, nor necessarily agree, or disagree with the paper's conclusions; rather, I will remark on two unmentioned reasons that I suggest underwrite some of the resistance to the *very idea that human beings could, in actuality, be simulatable, if we define simulation as our development of artificial entity(s) that operationally achieve sociocognitive parity with humans*. The first is the threat that such a possibility poses to two of the most generic, identitarian properties of personhood – the belief in our substantive, autonomous identity as persons possessing the property of being the person whom we experience as "me" or "I." The second is "our" autocentric relationship to other forms of living, and *emergent* entities.

"Operational Autonomy," "Being No One" and the Principle of Mediocrity

Imagine a given person *x,* negotiating the contingencies of *a posteriori* experience – what Alan Turing called the "Informality of Behavior." That is, they live one day out of all the days that constitutes their life; a day being normatively defined as a 24-hour day starting at 12:01:01 AM and ending at 11:59:59 PM. Formally, they have to negotiate a given set $B \supseteq\supseteq (A \cup A^{-})$ where B is the subset of all possible decision scenarios that must be made on, for example Thursday, March 27, 2007, and where $(A \cup A^{-})$ represents the totality of all possible implemented and unimplemented decisions. Furthermore, whatever decisions are decided and acted upon by *x,* these decisions will be a function of: one, intentionalized, and, non-intentionalized (unconscious) mental processes and two, third person, contingent events occurring during Thursday.[11] Therefore, given that $(A \cup A -)$ delineates *N* possible epistemic alternatives, B represents a potentially infinite set of alternatives.

We certainly have at our disposal a number of meta-analytical frameworks to select from as we attempt to understand how *x* negotiates the day of Thursday, March 27, 2007. For example Michel Foucault's work provides a potentially very important framework because his radical nominalism underscored the *a priori,* interlocking (theoretical) constraints of discursive "formations," "enunciative fields" and the ever present movement of power relations – the "micro-physics of power"[12] – each of which provides explanatory support for a constructionist, socio-political understanding of a history that both enables and constrains human "operational autonomy."

What certainly would not be present is any appeal to the traditional epistemic descriptions of phenomenological experientiality, empiricist theories of referentiality, or the naive realism of the folk psychological categories of "personality," "emotions," and "learning" deployed by Yilmaz, Ören and Aghaee. Furthermore, in Foucault's case, the use by each person of the term "I" has no function other than its indexical, enunciative role as nominative and objective personal pronouns spread out on the plane of discursivity and power relations. Notwithstanding its genuine explanatory potential, a theory like Foucault's cannot satisfactorily take into account any *every*day situation where each human person must be functionally autonomous. Foucault's formulation of the concept of freedom with respect to human action is simply not plausible as it stands.[13]

Any theory of "operational autonomy" must envision modes of autonomous human action that take into account the theoretical pragmatics of decision procedures that our everyday lives require for us to function in the first place. Furthermore, Foucault certainly would not have accepted the commensurable position that human discursive, socio-political, and quotidian history are embedded in, and constrained by the evolutionary history of a specific anthropic entity referenced by the species-determinate class "human beings(s)."[14]

Notwithstanding these points of dispute concerning positions like Foucault's, there is a significant point of intersection between his nominalist critique of the Cartesian-Kantian, autocentric tradition of the transcendental "I," and, one of the most interesting neuro-philosophical bodies of work to emerge in recent years on the question of consciousness and self-consciousness. Thomas Metzinger, Director of the Theoretical Philosophy Group at the Department of Philosophy of the Johannes Gutenberg-Universität, Mainz, has remarked, "The 'phenomenal first-person perspective' is one of the most fascinating natural phenomena we know, and in a certain sense we *are* this phenomenon ourselves" (Metzinger 2004a).

In describing consciousness as a *natural* phenomenon, care must be taken in how we are too interpret this phrase: it is quite easy to infer a folk psychological reading that references an experience to which we attribute existential substantiveness. Like Foucault, Metzinger denies that any naïve readings of "things" indexed by the personal pronouns "I," "you," "we," "she," or "him" have any ontological status. As the title of his most recent monograph makes clear, "we" are "No One" (Metzinger 2004b).[15] Metzinger describes his approach theoretically as PSM (the "phenomenal Self-Model): a "virtual" "simulation" or "avatar" that is systemically generated at the operative level of the neurobiological, neurocomputational, and functionalist activity of the brain. As a "pre-reflective," evolutionary development, this "episodically active representational entity" permits "us" to experience the world phenomenally as a non-representational, immersively transparent experience. Accordingly, it generates an illusionary (virtual) experience of a "singular, temporally extended experiential self" functioning in the world.[16]

Metzinger argues that we can never, under "traditional" folk and philosophical conditions, convert acts of self-reflexive introspection into meta-reflective acts of self reflection *on* first person consciousness, insofar as its nature is "nonconceptual," "subdoxastic" and "phenomenally transparent." Consequently, "we" are, by definition, always fundamentally in error with regards to *some* aspect about our(selves) and the world. Metzinger states:

> What makes a phenomenal representation transparent is the attentional unavailability of earlier processing stages in the brain for introspection. The instruments of representation themselves cannot be represented as such, and hence the system making the experience, on this level and by conceptual necessity, is entangled in a naïve realism: In standard configurations, one's phenomenal experience has an untranscendably realistic character. (Metzinger 2004b.)[17]

Setting aside the issue of whether Metzinger's arguments are sustainable as philosophical proposals, his work is now playing a major theoretical role in the "Cronos Project" that is being funded by the Engineering and Research Council (UK).[18] Headed by Professors Owen Holland and Tom Troscianko, the project has an "adventurous" research objective: it seeks to develop the first robotic entity to display Strong Machinic Consciousness (SMC). Whether this project will fail or not is, for me, not the question that must be addressed since there is a far more profound issue that we all may have to face. The history of (S)AI research is riddled with repeated failure that has become one of a number of contributing factors fostering the continued presence of an autocentric belief in our own uniqueness. Granting the fact that from an evolutionary standpoint "we" are extremely rare, this fact can *not* then be used to support the stronger claim that "we" are absolutely unique.

Conclusion

As Carter realized, the Generalized Copernican Cosmological Principle was having a detrimental effect on cosmological research, hence his counter proposal of the Anthropic Principle. It is important to understand that this did not sideline the Principle of Mediocrity. The point of the Anthropic Principle is that it was not proposed to establish our uniqueness, but rather to note that, as a particular type of evolutionary, Anthropic entity, we must take into account the selection effects being human has on research. One interpretation of the Mediocrity Principle is that there is no privileged Anthropic entity *x,*

such that *x* can be privileged over Anthropic entity *y*. I strongly suggest that what we may, conditions permitting, have to confront critically is the emergence of new forms of anthropic entities, be they robotic, or, as Yilmaz, Ören and Aghaee note, simulable "agent societies (that) include being rational, responsible . . .accountable . . . trustworthy . . . an ethical agent, moral agent."

I recommend that anyone who maintains a strongly skeptical position with regard to the arguments of Yilmaz, Ören and Aghaee read the 2006 report, "Roboethics Roadmap," issued by the European Robotics Research Network from which, to finish, I quote:

> In terms of scope, we have taken into consideration – from the point of view of the ethical issues connected to Robotics – a temporal range of a decade, in whose frame we could reasonably locate and infer – on the basis of the current State-of-the-Art in Robotics – certain foreseeable developments in the field. For this reason, we consider premature – and have only hinted at – problems inherent in the possible emergence of human functions in the robot: like consciousness, free will, self-consciousness, sense of dignity, emotions, and so on. Consequently, this is why we have not examined problems – debated in literature – like the need not to consider robots as our slaves, or the need to guarantee them the same respect, rights and dignity we owe to human workers. Likewise, and for the same reasons, the target of this Roadmap is not the robot and its artificial ethics, but the human ethics of the robot's designers, manufacturers and users. (Veruggio 2006: 22.)
>
> It is the first time in history that humanity is approaching the threshold of replicating an intelligent and autonomous enitity. This compels the scientific community to examine closely the very concept of intelligence – in humans, animals, and of the machines – from a cybernetic standpoint. (Veruggio 2006: 7 and 22.)[19]

Notes

---

[1] S(W)AI, or the principles of Strong and Weak Artificial Intelligence, were first proposed by the American philosopher John Searle to distinguish between Weak AI and Strong AI which, in his seminal 1980 paper "Minds, Brains, and Programs," he described as follows: "according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind."

[2] This was also the first time that the phrase "Artificial Intelligence" appeared in print. However, neither McCarthy nor Alan Turing was responsible for the birth of the neural computational thesis. It was first presented in the paper by McCulloch and Pitts, "A logical calculus of the ideas immanent in nervous activity" (1943).

[3] Dirac (1974) states: "The Large Numbers hypothesis asserts that all the large dimensionless numbers occurring in Nature are connected with the present epoch, expressed in atomic units, and thus vary with time. It requires that the gravitational constant *G* shall vary, and also that there shall be continuous creation of matter. The consistent following out of the hypothesis leads to the possibility of only two cosmological models. One of them, which occurs if one assumes that the continuous creation is a multiplication of existing matter, is Einstein's cylindrical closed Universe. The other, which occurs if one assumes the continuous creation takes place uniformly through the whole of space, involves an approximately flat Minkowski space with a point of origin where the Big Bang occurred."

[4] [4] The Russian physicist and mathematician Abraham Zelmanov was the first to propose a modernist version of the argument from anthropic experience (see Rabounski 2006). The Serbian physicist Milan

Cirkovic traces the conceptual framework for anthropic arguments back to the work of Empedocles of Acragas "(sixth to fifth century BC) " (see Cirkovic 2002).

[5] There is little doubt that Carter's Bayesian formulation was drawn from Bostrom's work given the citation of Bostrom's work (see Bostrom 2005).

[6] The SSA must be premised, a priori, on the existence of an observer who is, by definition, epistemologically-constrained. Hence, Bostrom's important addition of SSA to Bayesian analysis. See the following section on Descartes for a detailed elaboration of the problematics surrounding the role of the observer.

[7] Michael Hanby notes that "Descartes' fundamental principle negates the traditional God only to reconstruct him as a causal hypothesis and guarantor of clear and distinct ideas."

[8] It is important to reiterate the foundational role that Descartes gives to his reconstructionist program through his pivoting of the project on first person, epistemic experientiality – each individual person's first hand, or "first person" interaction with the world. Thus anyone who initiates an epistemological investigation of the world must, by definition, commence from their individuated cognitive activity, and, there must, by definition, exist the philosophical possibility that, in the face of the methodological application of systemic skepticism, there are object(s) or event structure(s) in the world that are the cause of experiences in the first place. And further, they must presume that they, as humans, are capable of rationally substantiating their experiences of this "world."

[9] The most succinct contemporary formulation of the "problem" is given in (Chalmers 2004): "The task of a science of consciousness, as I see it, to systematically integrate two key classes of data into a scientific framework: third-person data, or data about behavior and brain processes, and first-person data, or data about subjective experience. When a conscious system is observed from the third-person point of view, a range of specific behavioral and neural phenomena present themselves. When a conscious system is observed from the first-person point of view, a range of specific subjective phenomena present themselves. Both sorts of phenomena have the status of data for a science of consciousness."

[10] An extremely influential variant of this tradition is represented by the work of Michel Foucault. The major difference is, of course, Foucault's rather behaviorist principle of "exteriority" which jettisons the Cartesian theory of mind. As is well known, Foucault argues that we do not have a "mind" at all.

[11] For example, tripping over an ant. No matter how remote, or absurd, the possibility of someone actually tripping over an ant is, this scenario is logically possible. One cannot, in the face of the principle of contingency, assert a priori what will, or will not occur on that Thursday.

[12] The phrase "micro-physics of power" was first used by Foucault in *Discipline and Punish: The Birth of the Prison* (English translation 1977). This is one of a complex series of second-order explanatory concepts that Foucault used to explicate the inter-relationship between persons, as **discursive-physicalistic** entities, and the pragmatics (political) power as a "micro" localized, historical process.

[13] This is hardly surprising, given Foucault's longstanding, critical interest in Kant's anthropology. We should remember that Foucault's 1961 doctoral dissertation, "Introduction to Kant's anthropology from a pragmatic point of view," was supervised by the French philosopher and physician Georges Canguilhem, one of the most important, and influential, neo-Nietzschean philosophers of science, and, most crucially, of biology, in the twentieth century.

[14] Foucault's work is not the only theoretical model for articulating a structural relationship between Metzinger's proposals concerning the primacy of phenomenal experientiality and the materiality of a history that is embodied in discursive, institutional, and productive practices that also frame human action. Clearly, the work done, and being done in the cognitive sciences, analytic philosophy, and the neurosciences is of enormous importance. There is, though, a poor, if not non-existent, understanding of the constitutively functional role of the complexities of the various, and, at times, incommensurable histories that are also evolutionary determinants in the emergence and ongoing socio-cognitive life of humanity. What Foucault's work does provide is one alternative to understanding how these "histories" operate with respect to the issue of cognition and the world. A prime indicator of this problem is the ever-present referencing in Metzinger's work of the inter-locking precepts of folk psychology and naive realism. As discussed later in this paper, Metzinger is quite cognizant of the problematics of "getting around" these precepts if his arguments are to be sustained. The difficulty is that the precepts mask a highly problematic form of explanatory reductionism when it comes to understanding the mediative force of the discourses, institutions, and the "micro-physics of power" in the life of humans. It is clear that a great deal of the literature that Metzinger is steeped in views this "force" as either non-existent or a mere "naïve" "folk" add-on, or worse, as an epiphenomenon of the neurocomputational, informatic and neurobiological functionality of the brain, rather than as the determinative constituents of our understanding of the nature of mind and its interface with the world.

[15] This rather dramatic title has led itself to some misunderstanding of Metzinger's work. He himself has stated that the title of an earlier work – *Subject and Self-Model: The Perspectivalness of Phenomenal Consciousness Against the Background of a Naturalist Theory of Mental Representation* – is better suited as a description of the content of *Being No One.*

[16] The distinction between consciousness and self-consciousness is essential to Metzinger's theory. Phenomenal experientiality is comprised of immersive consciousness (PSM) and, self-consciousness as directed intentionality, what he refers to as the "Phenomenal Model of the Intentionality Relation (PMIR)."

[17] There is little question that attempting to formulate a theoretical, naturalistic (evolutionary) model of first person consciousness leads to some very difficult epistemological and methodological problems for Metzinger. As he states, "In standard configurations, one's phenomenal experience has an untranscendably realistic character." The key phrase here is "In standard configurations." It is important to note that Metzinger explicitly locates his work within the historical context of the objectives of the high, eighteenth-century scientific Enlightenment. In particular, this involves an historical, if not utopian, vision of the progressive amelioration and development of the human condition through the accumulation of scientific knowledge. But, as Metzinger himself noted with regard to a foundational, initial point for some of the philosophical framework for the Enlightenment, the Cartesian program for the revealing, in its full transparency, an ontology of the content, if not nature, of self conscious identity – such a project is always "entangled in a naïve realism." The issue is how Metzinger resolves the problem of disentangling "naïve realism" from his own theoretical model for phenomenal consciousness. His response is implied by the phrase "In standard configurations," by which he refers to traditional practices that have been endemic to philosophy such as phenomenological, discursive, logical, and, of course, metaphysical analysis. Metzinger first, and foremost, structurally harnesses these traditions directly to the rapidly emergent, post-war disciplines of the cognitive and neurosciences, and to philosophies of mind strongly constructed around computationalist-functionalist and neural-computationalist explanatory models. While there is nothing new in this, he has also, more radically, immersed himself in the research areas of psychiatry, neuro-psychiatry, and neurology in order to develop and test his PSM model, based on the methodological and operational premise that one can reverse engineer his models in light of experimental work into the etiology of a number of severe cognitive and psychiatric disorders where our normative

personal, inter-personal, and inter world modes of cognition have been seriously disrupted, or worse, completely collapsed.

[18] However, I recently met Dr. Holland at a conference in Prague where he told me that the project is on hold because his present funding has run out.

[19] Notwithstanding Dr. Veruggio's predictive remarks concerning humanity's "approaching the threshold of replicating an intelligent and autonomous enitity," this observation cannot be taken a face value. As I have already mentioned, the history of (S)AI claims and predictions is littered with failure, backpedaling, and programmatic restructuring. Whether these entities will actually emerge depends on the resolution of a number of extremely complex scientific and philosophical issues, and must be also weighed against a host of economic, socio-political, ethical, and ecological factors that could militate against their successful development. In the section of his book titled "Is artificial subjectivity possible? Could there be non-biological phenomenal selves?", Metzinger considers in some detail the theoretical and methodological constraints that would have to be in place to achieve, or not achieve, SMC. During this review, he addresses what he refers to as the "Big Technological Dream" (620). In his discussion of the possibility of the successful implementation of SMC, at least in the short term, he also underscores emerging work in the research areas of "cognitive robotics" and "hybrid biorobotics" that may eventually be successful. Setting aside the question as to whether these efforts will achieve success, Metzinger, more importantly, asks why we should even consider going forward with this research: "Put very simply, we would dramatically increase the amount of suffering, misery and confusion on the planet . . . it is not at all clear if the biological form of consciousness, as a desirable form of experience, an actual good in itself, something that one should keep on multiplying without further thought" (620) … "Phenomenal experience is not something to be unconditionally glorified. Among many other new properties, biological self-consciousness has brought an enormous amount of misery … into the world, an ocean of phenomenal suffering, which was simply not there before. As one of my students once put it: 'The universe may be a good place for evolution, but not such a good place for individuals'" (623).

References

Bostrom, N. 2002. *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. New York: Routledge.

Bostrom, N. 2003. Are You Living in a Computer Simulation? *Philosophical Quarterly* 53: 243-55.

Bostrom, N. 2005. Self Location and Selection Effects: An Advanced Introduction. URL www.nickbostrom.com.

Carter, B. 1974. Large Number Coincidences and the Anthropic Principle in Cosmology. In *Confrontation of Cosmological Theory with Astronomical Data*. Proceedings of the Symposium, Krakow, Poland, September 10-12, 1973. Ed M. S. Longair. (A75-21826 08-90) Dordrecht: D. Reidel Publishing Co.: 291-98.

Carter, B. 2004. Cosmology: Facts and Problems. Colloquium, College de France.

Carter, B. and McCrea, W.H. 1983. The Anthropic Principle and its Implications for Biological Evolution [and Discussion]. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 310(1512): The Constants of Physics (Dec 20, 1983): 347-63.

Chalmers, D. 1995. Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies* 2 (3): 200-219.

Chalmers, D. 2004. How Can We Construct a Science of Consciousness? In Michael S. Gazzaniga, ed., *The Cognitive Neurosciences III*. (Cambridge, MA and London: MIT Press): 1111-1121.

Cirkovic M. 2002. Ancient Origins of the Modern Anthropic Cosmological Argument. *Astronomical and Astrophysical Transactions,* Vol. 22, No. 6, December 2003, pp. 879–886.

Descartes, R. 1901. Meditations on First Philosophy. John Veitch translation of 1901. Home page Descartes' *Meditations*. URL http://www.wright.edu/cola/ descartes/ Meditations II, 3.

Dickie, R. H. 1961. Dirac's Cosmology and Mach's Principle. *Nature* 192: 440-41.

Dirac, P. A. M. 1937. A New Basis for Cosmology. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 165, Issue 921: 199-208.

Dirac, P. A. M. 1974 Cosmological Models and the Large Numbers Hypothesis *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences,* Vol. 338, No. 1615 (Jul. 16), pp. 439-446.

Dreyfus, H. L. and Paul Rabinow. 1983. *Michel Foucault: Beyond Structuralism and Hermeneutics*. Second edition. Chicago: University of Chicago Press.

Foucault, M. 1977. *Discipline and Punish: The Birth of the Prison*. Translated by Alan Sheridan, London: Allen Lane, Penguin. First published in French as *Surveiller et punir* (Paris: Gallimard, 1975).

Graham-Rowe, D. 2005. Mission to build a simulated brain begins. *New Scientist*. June. Retrieved from URL http://www.newscientist.com/article.ns?id=dn7470.

Hanby, M. 2003. *Augustine and Modernity*. London: Routledge.

McCarthy, J., Minsky, M. L., Rochester, N. and Shannon, C. E. 1955. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. John McCarthy, Department of Mathematics, Dartmouth College, Hanover, New Hampshire.

McCulloch, W. S. and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* 5: 115-33.

Metzinger, T. 2004a. The subjectivity of subjective experience: A representationalist analysis of the first-person perspective. Revised version in *Networks* 3-4: 33-64.

Metzinger, T. 2004b. *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

Newman, L. 2005. Descartes' Epistemology, *Stanford Encyclopedia of Philosophy*, URL http://plato.stanford.edu/.

Rabounski, D. 2006. Zelmanov's Anthropic Principle and the Infinite Relativity Principle. *Progress in Physics* 1: 35-37.

Rudnicki, K. 1995. *The Cosmological Principles*. Cracow: Jagiellonian University.

Turing, A. M. 1950. Computing Machinery and Intelligence. *Mind* 59: 433-60.

Veruggio, G. (2006). Euron Roboethics Roadmap. Euron Roboethics Atelier Genoa, 27 February - 3 March.

Yilmaz, L., Ören, T., Aghaee, N-G. 2006. Intelligent Agents, Simulation, and Gaming. In *Simulation & Gaming* 37: 339-49.

Zelmanov A. L. 1944. Chronometric Invariants. Dissertation. First published: CERN, EXT-2004-117, 236 pages.