Human Version 2.0
Between the "Banality" and
"*Ressentiment*" of Neuroengineers

Norman K. Swazo

Abstract

Neuroengineers such as Hugo de Garis propose to build artificial brains called "artilects" that go beyond the cybernetic model seeking to build "cyborgs." Empirical research in this field raises moral questions to which both moral psychology and moral philosophy can contribute, especially to analysis of the motivations of a neuroengineering project such as that of de Garis. The problem is that while empirical results of recent moral psychology provide us with explanations and interpretations having empirical probability, they remain descriptive without resolving moral implications in the sense of having prescriptions on the central question whether the neuroengineering project should be pursued. Assuming neuroengineers can deliver on their projects by the mid- to late- 21st century, both the neuroscience community and philosophers have at least an epistemic (if not a moral) responsibility to engage the moral dimension of this scientific and technological quest.

Key Words: artificial intelligence; artilects; neuroengineering; morality; Hugo de Garis

Half the harm done in this world
Is due to people who want to feel important
They don't mean to do harm—
But the harm does not interest them.
Or they do not see it, or they justify it
Because they are absorbed in the endless struggle
To think well of themselves.
T. S. Eliot

Introduction

For many centuries, from the time of Plato through to the 19th century philosophy of Friedrich Nietzsche, the contest of rationality and animality in the human being has been the center of its self-determining agon, expressed as the tension between the rule of reason and the inclinations of emotion. In the 20th century, the German philosopher Martin Heidegger wrote that the unity of rationality and animality remain as yet *undetermined*, even as he warned of the need for a different conception of human being that abandons this metaphysical duality. Today, however, the battle between rationality and animality in the human being takes center stage not only in empirical research pursued by moral psychology, the latter distinguishing itself methodologically from moral philosophy, but also in the promises of neuroscience and neuroengineering.

Corresponding author: Norman K Swazo
Address: Professor of Philosophy, University of Alaska Fairbanks, USA
e-mail: ffnks@uaf.edu

It is precisely this distinction of methodological commitments that contributes to sorting out moral implications of contemporary neuroscience and what is now known as neuroengineering, such as is pursued by the "brain builder" Hugo de Garis. His is the kind of research towards prospective brain building that raises important moral questions: What does one say, for example, in an encounter of deeds seemingly monstrous in their conception or in their execution, when the doer of the deed is himself said to be neither demonic nor monstrous, *nicht einmal unheimlich*— "not even sinister," as the philosopher Hannah Arendt would put? (Arendt, 2006). That is a question one can pose with reference to de Garis's work.   Arendt understood the power of *thoughtlessness*, to be distinguished from the power of *knowledge*, where one encounters a phenomenon amidst human life she termed "the banality of evil," manifest by a "shallowness in the doer that [makes] it impossible to trace the incontestable evil of his deeds to any deeper level of roots *or motives.*" (Bernstein, 1996)   The neuroengineering project as conceived by de Garis proceeds such that Arendt's concern for thoughtlessness remains a prominent concern for philosophers pressed to discern and articulate the moral implications of this kind of research.

At the same time, of course, one can question the legitimacy of any human perspective on life that is anthropocentric, especially when defensive, in the sense of presumptuously safeguarding and privileging the human being over and above all evolving forms of life—including forms of life conceivably to be engineered by the very power of science. In today's disputations this includes developments in human genomics, surely, with the prospect of human cloning. But it also includes research that advances a cybernetic model of the human being, thereby forecasting the human life-world in the context of cybernetic architecture. One who would question the validity of anthropocentrism cannot but attend to the value of the whole of nature as we know it today, to speak then of nature's intrinsic value—without construing nature as a collection of things subject to the indiscriminate disposition of human beings. But, one who would challenge the anthropocentric perspective would also lend validity to what neuroengineers see as the prospect of a "cyborg" (cybernetic organism) species "created" consequent to the successful merger of neuroscience and quantum physics. In such a "species" the human *integrates* with machine, thus ostensibly to form the supposed next step in human evolution, even to the point that the human then becomes "sub-species" to this new "life-form" (also called "nano-teched matter"). In this cybernetic conception, we may suppose, "the rational" seeks it final dominance over "the animal" that is in us, as artifice displaces the biological.

But it can also be said, with good reason, that one has here (in a word): *subterfuge*—a deception by artifice. There is deception insofar as precisely what is proposed to be *in* the human interest is already, in the very initiative, *against* the human extant today as *homo sapiens sapiens*. The goal is to displace this species with the creation of a superior, dominant artifice. A philosopher cannot but ask: Is there evil here, in the sense of either the monstrous or demonic, or even the banal, such as Arendt conceived it in speaking of the banality of evil?   It is this question that motivates this commentary.

Human Version 2.0: The "Artilect" Prospect

In recent time the BBC *Science* television channel presented a program entitled, "Human Version 2.0," describing events in which neuroscience and biotechnology combine to produce a moment called "the singularity"— which is when the human intellect in effect creates non-biological intelligence that either integrates with human intelligence or exists independently as self-enhancing artificial intelligence.   There is more than mere integration at work here, as the concept 'cyborg' already suggests: Here, in fact, is a scientific quest to see the human being as creator, indeed as a species, superseded, displaced, and even replaced, by its creation.

de Garis speaks of an emergent entity, to be created by neuroengineering, which he

terms an *artilect*. This word stands for "artificial intellect" and represents the prospect of an artificial brain with a function that exceeds—"by a factor of trillions of trillions of trillions of times"— what is now known of human brain function, i.e., its computational ability. Why such a multiplier effect? de Garis answers:

> …the size of the electronic components in mass memory chips will be such that it will be possible to store *a single bit of information* on a *single atom*. (A "bit" is a "binary digit," a 0 or a 1, that computers use to represent numbers and symbols to perform their calculations.) So how many atoms (and hence how many stored bits) are there in a human sized object, such as an apple? The answer is astonishing—a trillion trillion atoms (bits), i.e., a 1 followed by 24 zeros, or a million million million.

> Are you beginning to get an inkling for why I believe that massively intelligent machines could become trillions of trillions of times smarter than we are later this century? (de Garis, 2005).

Such is the horizon of "quantum computing." de Garis claims to be "a social critic with a political and ethical conscience" despite his prominent position in research designing and building "artificial brains." He has stated: "I am very worried that in the second half of our new century, the consequences of the kind of work that I do may have such a negative impact upon humanity that I truly fear for the future." (de Garis, 2005). He adds, however: "The prospect of building godlike creatures fills me with a sense of religious awe that goes to the very depth of my soul and motivates me powerfully to continue, despite the possible horrible negative consequences." Thus, de Garis admits to feeling "quite 'schizophrenic' about this."

Clearly, de Garis points to what many may consider an ominous future for humanity if he and other neuroscientists or

neuroengineers are successful with what they propose and foresee. At the heart of this research is a fundamental ethical dilemma which science itself cannot and, so, does not answer. The dilemma is not to be resolved by appeal to the authority of any empirical method, though empirical facts inform the moral deliberation. The dilemma is present in a researcher such as de Garis himself, in what he terms his schizophrenia. But, this dilemma may also be represented as a dispute between de Garis's manifest *scientific understanding* and his ambivalently engaged *moral imperatives*.

Yet, there is also an admitted lack of understanding; for, he admits, "evolutionary engineering" such as he promotes means "it is possible to build something successfully without understanding fully how it functions." One might conclude that ignorance rather than understanding prevails, given de Garis's insistence on continuing with the research despite his moral qualms. We find here an insistence that manifests itself as an egocentric comportment, by no means articulated as an appeal to epistemic duty. de Garis is, to use his own classification scheme, a "cosmist," i.e., one who favors building artilects: Building artilects is for him "like a religion…" (de Garis, 2005). "Cosmism," he opines, " is a 'scientist's religion,' i.e., a religion in the sense that traditional religions energize and direct the lives of groups of people." (de Garis, 1996) Both his belief and action are given their motivation thereby: "If you are a Cosmist," he claims, "you will place higher priority on the creation of godlike artilects than the survival of the human species." (de Garis, 2007) Only such "new gods" are likely to be "more capable than human beings of answering the big questions" that have "cosmic significance." Seeing himself as one having both "high intelligence" and "visionary creativity," whose work is not to be dismissed as mere "science fiction," de Garis finds "the lack of vision of the majority of the highly intelligent people irksome." The fact is, dissatisfied with "the human brain with its pitiful trillion neurons" and "massively inferior memory capacities," de Garis has an axe to grind, and the existence of the artilect is his primary *raison d'etre.*

de Garis anticipates an objection to his forecast for the 21st century. Someone may

object to his "assuming implicitly that massive memory capacities and astronomical computational capacities are sufficient to generate massively intelligent machines…" He recognizes some researchers will claim, "there is more to producing an intelligent machine than just massive computational abilities," e.g., the presence of consciousness, which (to meet the objection) would then entail "creation of artificial consciousness in machines." Such a prospect is accepted by de Garis: "If science wants to build an intelligent conscious machine, then one obvious strategy is to copy nature's approach as closely as possible." (de Garis, 2005). Moving in this direction, de Garis prefers employing "molecular scale tools (called "nanotech" or "nanotechnology")," so that researchers such as he can take up "established neuro-scientific principles and incorporate them rapidly into our artificial brain architectures." (de Garis, 2005) Whatever the objection, the scientific quest as de Garis conceives it faces no unsuperable problem. The neuroengineers can get it right, sooner or later.

As always, the problem with frontier-edged research is the prospect of military expropriation, which de Garis sees forthcoming in the production of "intelligent soldier robots and other artificial brain based defense systems." (de Garis, 2005). The frightful questions emerge: "If they [artificial, conscious robots] become a lot smarter than human beings, might they decide that humans are a pest, a cancer on the surface of the planet, and decide to wipe us out?" "Should humanity take the risk, that that might happen?" Curiously, and ominously, de Garis conceives of his own place in this development by analogy with those involved in the Manhattan Project who created the atomic bomb. He realizes his own problem of conscience as artilect builder, his work "a stepping stone" to the "mass horror" he forecasts. But he persists in his quest: His will to power governs and overrules his moral sensibility—in part.

There is a problem much more fundamental than de Garis and company

appreciate. It is a problem that Heidegger (1972) warned about in holding that the question before us is not merely "What ought we to do?" but, rather, "*How must we think?*" The problem of human thinking is not a problem of human knowledge, in the classical sense of sorting out the possibilities of a theoretical or scientific knowledge of which humans are capable and then facilitating its responsible progress in technological innovation and advancement that serves the human interest in globally diverse ways. It is the problem Heidegger characterized in his concern for the totalizing installation of the technological, expressing itself in the human life-world through what he called "*rechnendes Denken*," i.e., *calculative* thinking. This is a kind of thinking that stands in contrast to *besinnliches Denken*, i.e., *reflective* or *meditative* thinking.

The very quest for creation of artilects and cyborg species rests upon a neuroengineering commitment to explore the limits of calculative thinking. Its goal is to realize a quantum-level of computing power to be carried out ostensibly *in* the human interest but also *against* the human interest. The goal is conceived to be in the human interest insofar as artificial intelligence enhances human creative power and facilitates human productivity. The goal is against the human interest insofar as neuroengineers expect (for them, quite tolerably) the human species to be superseded by the inevitably dominant species the concept 'cyborg' represents. And even then, beyond the very possibility of the cyborg, which would retain elements of human biology in its integrated state, is the very idea of an artilect that is entirely machine, conscious and even self-conscious of its quantum-level computational power, eventually capable of advancing its own interest. This entity is for de Garis "god-like" precisely because such machines can perform in "god-like" ways, their productivity even continuing the very cosmic process of creating universes (which, lest one think oneself here yet in the realm of science fiction rather than scientific fact, for this is already a theoretical, mathematically calculable, reality for quantum physicists).

The Moral Dilemma

Let us consider de Garis' engagement of the moral dilemma as he presents it in a paper published in 1990. (de Garis, 1990) de Garis calls for "a new branch of applied moral philosophy" that can engage the ethical issues associated with creation of "advanced artilects" having "unfathomable complexity and possible 'Darwinian' self modification." One immediate problem is that de Garis does not engage philosophical analyses such as those published by John Searle, D.C. Dennet, or Hubert Dreyfus. Nonetheless, de Garis premises his fears upon the differences between human "fixed architecture" in which "human neurons communicate at hundreds of meters per second" and that of artilects in which "electronic components communicate near the speed of light." (de Garis, 1990) It is this speed of communication and self-enhancement capability that has de Garis focusing on the Darwinian element as central to the moral question.

He considers "human moral attitudes" to be "psychological attitudes"—"ultimately physical/chemical phenomena"—consistent with operative assumptions in neuroscience. "What human beings consider as moral," he adds, "is merely the result of our biological evolution." Given these premises, de Garis asks: "Could one not argue that mankind has not only the power, but the moral duty to initiate the next major phase in evolution, and that it would be a 'crime' on a universal or cosmic scale not to exercise that power?" (de Garis, 1990).

That is where de Garis leaves it— he provides no moral reasoning one way or the other in this paper in answer to his question. But, other commentaries he has provided make it clear he answers the question in the affirmative: (1) Assuming artificial intelligence scientists/neuroengineers have the power to construct artilects, such scientists/neuroengineers have a moral duty to initiate the next phase in evolution; and (2) it would be a crime on a universal or cosmic scale for AI scientists/neuroengineers not to exercise

that power. Thus, de Garis's moral position has a twofold character. On the one hand, he is a reductionist insofar as he reduces morality to biological (neuro-chemical) processes, in which case an answer to his question is to follow from sociobiology or moral psychology; on the other hand, he is deontological insofar as he stipulates a moral duty for this segment of the scientific community as representatives of humanity, in which case an answer to his question may be engaged via deontological ethical theory. Accordingly, one can evaluate de Garis's position on the basis of these two theoretical domains.

A. Moral Psychology

Consider some prominent empirical work in moral psychology, such as that of Jonathan Haidt and Joshua Greene. Haidt holds that humans who are moral have "social motivations beyond direct self-interest." (Haidt, 2007; Haidt, 2001) Distinguishing between "affect" and "cognition," he informs us, "Evolutionary approaches to morality generally suggest affective primacy"—that is, the primacy of emotion over reason, such that "higher-level human thinking is preceded, permeated, and influenced by affective reactions (simple feelings of like and dislike) which push us gently (or not so gently) toward approach or avoidance." Haidt prefers to speak of "moral intuition" as "fast, automatic, and (usually) affect-laden processes in which evaluative feeling of good-bad or like-dislike (about the actions or character of a person) appears in consciousness without any awareness of having gone through steps of search, weighing evidence, or inferring a conclusion."

Haidt argues further, also on appeal to empirical evidence, that "When we engage in moral reasoning, we are using relatively new cognitive machinery that was shaped by the adaptive pressures of life in a reputation-obsessed community." (Haidt, 2007) Thus, "in daily life the environment usually obliges us by triggering some affective response." In short, reputation and "indirect reciprocity" of relationship are important elements of moral response. But, Haidt also accounts for what he and other researchers call "five psychological foundations, each with a separate evolutionary origin, upon which human cultures construct

their moral communities." (Haidt, 2007). Liberally-minded individuals "value and rely upon the harm and fairness foundations primarily." But, beyond this set of concerns, others have "intuitions about ingroup-outgroup dynamics and the importance of loyalty" as well as "intuitions about authority and the importance of respect and obedience" and "intuitions about bodily and spiritual purity and the importance of living in a sanctified rather than a carnal way."

Assuming these claims to be empirically supported with some degree of epistemic reliability, de Garis's appeal to the authority of moral psychology is problematic on the research results advanced by Haidt. It is unclear that neuroengineered artilects would even have a structure/architecture that includes anything like affective primacy in relation to a *post hoc* moral reasoning. Quantum-level computing power such as de Garis envisions it would be developed entirely artificially, thus without biological (i.e., neuro-chemical) elements that are presumed to be the basis of affect or moral intuition. Only in the case of the integrated cyborg model would a "cyborg-artilect" have such a presumed affective contribution/basis to moral judgment. Consistent with current research (assuming these results remain valid), neuroengineers would have to preserve, in the integration of the cyborg-artilect, "areas of the medial prefrontal cortex, including ventro-medial prefrontal cortex and the medial frontal gyrus," since, Haidt says, "These areas appear to be crucial for integrating affect (including expectations of reward and punishment) into decisions and plans." (Haidt, 2007). Assuming some capacity for "internal conflict" in a cyborg, such as having to consider and determine whether a human being is either friend or enemy in a given situation, neuroengineers would be motivated to preserve in the integrated cyborg the function of "the anterior cingulate cortex, a brain region that responds to internal conflict." (Haidt, 2007).

Since, for researchers such as Haidt, "Moral reasoning, when it occurs, is usually a post-hoc process in which we search for evidence to support our initial intuitive reaction," (Haidt, 2007) a cybernetic approach to the production of a cyborg-artilect leaves open the question of "like/dislike" that manifests itself in a human as a distinction of friend/enemy. This prospective scenario, however, implies that cyborg-artilect "moral reasoning" would be *post hoc* and expressive of an initial affective response to any situation under review. Given the speed of quantum computation in the cyborg that becomes the analog of moral reasoning in the human, moral reasoning as such would be wholly irrelevant and the initial affective response in effect totally determinative of a cyborg-artilect's conduct. Any *post hoc* construction explaining or justifying a cyborg-artilect action would be quite beside the point except for a "community" of cyborg-artilects self-enhancing with a view to sustaining and improving upon group-loyalty and species dominance (all of this "grounded in reciprocity intuitions" currently recognized by moral psychology).

But there is a further point here. Haidt and colleagues argue that moral intuitions are *innately prepared* but generate *culturally variable* virtues. (Haidt and Joseph, 2004; Haidt and Graham, 2007; Haidt and Joseph, in P. Carruthers, *et. al.*, *in press*) They "believe that virtue theories are the most psychologically sound approach to morality" (Haidt and Joseph, 2004), though this is not to say "that virtue ethics is the best normative moral theory." Haidt concedes, he and his colleagues "speak only descriptively…" However, nothing here is straightforward or predictable: "different moral underpinnings provide the virtues with different eliciting conditions and different appropriate behaviors and responses." Presumably, the same would be expected of the integrated cyborg that retains key elements of biological matter in which these intuitions are elicited. Thus, the *cultural provenance* of the biological material used in the construction of cyborg-artilects can be a significant selection determinant of their behavior, consistent with the panoply of intuitions present or absent, given that "Western" and "traditional" societies

differ empirically on the moral intuitions manifest in human conduct.

What the foregoing tells us is that moral psychology can contribute to *description* and *prediction* of "moral" phenomena (assuming successful production of a cyborg-artilect), but moral psychology cannot tell us what *ought* to be the case, i.e., it cannot *prescribe* one way or the other on the question de Garis poses for applied ethics. It leaves unresolved the problem of how any neuroengineered artilect could or would have a *moral* computational capacity even as one may grant that an artilect would have a *nonmoral* computational capacity for any assortment of decisions that engage matters of fact without moral dilemma. This opens the door to de Garis's appeal to deontology.

B. The Deontological Evaluation

Immanuel Kant's moral philosophy is understood to be the principal modern formulation of deontological ethics. There is in each of us, says Kant, a moral law, which is what makes each human being a *personalitas moralis*, a moral person whose "dignity" elicits "respect" such that no human being is to be treated merely as a means to this or that end pursued by another, thus not to be treated as a "thing" subject to the conventions of extrinsic value manifest by way of "commercial value" (market price) or even "sentimental value," each of which changes according to some calculus of economic or personal interest. This moral law represents itself in imperatives of duty that have objective validity, and which hold true for rational beings as universal laws (akin to the objective validity of the universal laws of motion governing natural phenomena, such as were expounded by Isaac Newton in his *Principia*). Kant recognized that the imperatives of duty are readily challenged by the diversity of motivated human inclinations, such as those actions that follow from motives of self-interest. Nonetheless, for Kant competent moral reasoning must be self-consistent in the application of what he called "the

categorical imperative." Such reasoning would assure that otherwise merely subjective maxims (rules of action) would have their appropriate rational evaluation, so that human action could be guided by a motive of duty to the exclusion of mere inclination. Only objectively valid maxims that have the status of universal law to which rational beings would be subject could be determinative of human decision and so have moral authority.

Against Kant, moral psychologists appeal to evidence of neurobiological foundations of what Kant calls the moral law. They allow for "a theory of how the moral sense can be universal and variable at the same time." (Pinker, 2008) Joshua Greene, working in a direction much like Haidt, argues that "deontological philosophy, rather than being grounded in moral *reasoning*, is to a large extent an exercise in moral *rationalization*,"— quite in contrast to consequentialist ethics, says Greene, which (he argues) "arises from rather different psychological processes, ones that are more 'cognitive', and more likely to involve genuine moral reasoning." (Greene, "The Secret of Kant's Soul") To speak of moral rationalization in contrast to moral reasoning is, as psychologist Steven Pinker puts it, to "begin with the conclusion, coughed up by an unconscious emotion, and then work backward to a plausible conclusion." (Pinker, 2008). Indeed, Greene goes so far as to state he believes "it is possible that philosophers do not necessarily know what consequentialism and deontology really are." (Greene, "The Secret of Kant's Soul") But then, anyone may concur with this claim as formulated. Even a philosopher could readily accept that it is *possible* this is so, that philosophers do not *necessarily* know (and may not thereby have any appeal to necessity but only to defeasible, i.e., tentative, reformable, claims), and that they may not necessarily *know* (but only *believe*, with conditioned justification) what they assert about these two moral philosophies. Greene's statement, in short, is gratuitous. He leaves unexamined the varying implicatures present in concepts such as 'possibility', 'necessity', and 'knowledge', such that one may clarify the denotation, connotation, and domain of application of the concepts. Thus, there is no damage done to the authority of philosophical

pronouncement articulated by either deontologists or consequentialists, certainly not in the absence of adequate argument or admissible evidence to the contrary as normally expected in philosophical disputation.

Of course, Greene appeals to empirical evidence to assert deontology is merely a "philosophical manifestation" of a "psychological natural kind" (a "psychological pattern"), even a pattern "dissociable" from consequentialism as a different mode of moral thinking. What matters to him, then, are not the definitions in use as given by philosophers such as Kant, but the functional role of these patterns of thinking as this role can be garnered by empirical evidence. Greene's hypothesis is consistent with the position adduced by Haidt. Greene proposes: "If it turns out that characteristically deontological judgments are driven by emotion (an empirical possibility), then that raises the possibility that deontological *philosophy* is also driven by emotion (a further empirical possibility)…[Thus] what [deontological moral philosophy] really is, what it is *essentially*, is an attempt to produce rational justifications for emotionally-driven moral judgments, and not an attempt to reach moral conclusions on the basis of moral reasoning." (Greene, "The Secret of Kant's Soul")

Once again, as with the observations of Haidt, Greene's claims "are strictly empirical," though Greene holds, "if these empirical claims are true, [then] they may have normative implications, casting doubt on deontology as a school of normative moral thought." (Greene, "The Secret of Kant's Soul") This is not the place to dispute Haidt's or Greene's reading of Kant or subsequent developments in contemporary deontological ethics. For now, despite anticipation of some statement about normative implications of such research, we remain at the level of description. Thus, as Pinker (2008) put it, "Science amoralizes the world by seeking to understand phenomena rather than pass judgment on them." So it remains

with moral psychology as an empirical science concerned with the explanation and interpretation of matters of fact and which, presumably, is not thereby provided authority to commit the naturalistic fallacy of deducing ought from is. Thus, Pinker identifies one criticism of such research, viz., that, "the whole enterprise seems to be dragging us to an amoral nihilism, in which morality itself would be demoted from a transcendent principle to a figment of our neural circuitry." (Pinker 2008). But Pinker finds such a fear unwarranted. He allows for moral realism much like Plato did and much as mathematical realists understand the domain of numbers:

> …moral truths exist in some abstract Platonic realm, there for us to discover, perhaps in the same way that mathematical truths (according to most mathematicians) are there for us to discover. On this analogy, we are born with a rudimentary concept of number, but as soon as we build on it with formal mathematical reasoning, the nature of mathematical reality forces us to discover some truths and not others…Perhaps we are born with a rudimentary moral sense, and as soon as we build on it with moral reasoning, the nature of moral reality forces us to some conclusions but not others (Pinker, 2008).

Nonetheless, what matters in present commentary is what such empirical research seems to imply for de Garis's supposed moral concern for the project of neuroengineering artilects. If de Garis wishes to appeal to a notion of duty, to see the unrestrained and committed construction of artilects as a *moral* duty as well as a *scientific* duty, even Greene's empirical research displaces the validity of any appeal to deontology without first of all accounting for the psychological processes disclosed by such research. de Garis would have to recognize the affect-laden component of his own sense of moral ambiguity about his proposed neuroengineering of artilects. His call for applied thinking to this problem would itself then be merely *post hoc*, i.e., an effort to find *rationalization*, not *justification* (in the epistemological sense of the word), for

emotionally-driven affects he already has manifest in choosing to proceed with his neuroengineering project, despite the "horror" he forecasts.

C. Hail *Narcissus*

Pinker has allowed for the possibility of moral realism so that moral judgments may align properly with moral sense, with rationality itself entailing a manifest "external support for morality," viz., "that it cannot depend on the egocentric vantage point of the reasoner." (Pinker 2008) This observation points to yet another approach to moral analysis of the neuroengineering project. Whether one cares to accept the ascription or not, one finds in the work of a neuroengineer such as de Garis a singular expression of *narcissism*. Narcissism is represented to be "a pervasive pattern of grandiosity (in fantasy or behavior), need for admiration, and lack of empathy." It has its empirical pathology, of course, including "developmental short-circuiting" manifest as "affective deprivation," such that the narcissist manifests "defects of empathy." (American Psychiatric Association, 1994) Significantly, a narcissist can be caught up in self-contradiction; for, at the root of this narcissism is a "corrosive condition" known as *ressentiment*. There is in this *ressentiment* what E.M. Morelli argues is *ressentiment*'s essential element, viz., "a demand for and expectation of rational consistency," thus "a motive force." In a neuroengineer such as de Garis, *ressentiment* becomes creative of that which is explicitly *anti*-human precisely to safeguard what he takes to be fundamental to, yet manifestly lacking in, humanity as *homo sapiens sapiens*, viz., a self-consistent rationality eminently expressive of a calculative power such as he hopes will be realized in the engineering of quantum-level computation (Morelli, 1999).

This *ressentiment* is nothing temporary or fleeting in the way moods come over us and then pass. Rather, as Morelli says in reminder of the German phenomenologist Max Scheler's

clarification of this phenomenon, one can speak of *ressentiment* "as a permanent condition" of human existence. A neuroengineer who is also narcissist experiences the human condition as something nugatory, thus something to be resented, precisely because the relation of rationality and animality in the human is as yet undetermined though it "ought" to be determined in the direction of self-consistent rationality. In this dissatisfaction with the indeterminacy of *homo sapiens sapiens* the narcissist neuroengineer manifests a conjunction of elements—banality, calculative thinking, narcissism, *ressentiment*. It is this conjunction that impels the neuroengineering quest as pursued by men such as de Garis, and which adds to the complexity of the moral dilemma such as de Garis conceives it.

The ethical dilemma expresses itself thereby as a fundamental contradiction in the naricissist neuroengineer: between an *insistent self-assertion* that presumes humanity's entitlement to a supreme potency, on the one hand, and, on the other hand, a simultaneous yet *impotent hesitation* that discloses a "fundamental sense of insecurity and lack of self-worth." Resentment of the human as *homo sapiens sapiens* proceeds willy-nilly to *revenge against the human* in a fundamental antagonism. This antagonism is given voice in an utterance such as de Garis formulates, as in his suspect query: "Who is, what is to be the dominant species on this planet—human beings or the artilects?" That question has the structure of empirical prediction. It can be rephrased as a moral question, thus: *Should neuroengineers create an artilect, even as they cannot but forecast its dominance over the human species such as this species exists today?*

Even Nietzsche with his "hermeneutics of suspicion" of the rationalist quest understood that a destructive impulse (the "will to power") can be affirmed as a stimulant but nonetheless be appropriated in the service of that destruction. The "self-loving/self-loathing" *ressentiment* in de Garis is manifest in his statement:

> Ultimately, I think it would be a cosmic tragedy if humanity chooses to freeze evolution at the puny human level (with our pathetic little lives of 80 years in a universe billions of years old, that

contains a trillion, trillion stars--the "big picture"). For me, the tragedy of seeing the human species wiped out is less significant than not seeing the birth of the artilects. This sounds monstrous, and it is, in human terms, but to deny the creation of the first true artilect, which would be "worth" a trillion, trillion, trillion human beings, would be a far greater tragedy, a "cosmic" tragedy (de Garis, 2001).

With these words, the question posed at the outset of this commentary returns with urgency: What does one say in an encounter of deeds seemingly monstrous in their conception and in anticipation of the performance of the deed, when the doer of the deed is himself said to be neither demonic nor monstrous, *nicht einmal unheimlich*—"not even sinister," as the philosopher Arendt would put it? One can speak descriptively, even predictively, as a moral psychologist may. One can speak prescriptively, as a moral philosopher may, allowing the moral realist claim that despite culturally variable moral intuitions variability does not land us in an irresolvable rivalry (MacIntyre, 1989 and 2001). Yet, the prospective creations of neuroengineers such as de Garis call the scientific community to a more formidable question, which is itself not a moral question as such. Heidegger posed the question in his concern for "the planetary dominance of technology" and the gradual transformation of the human life-world into a technocratic order wherein human beings become mere things subject to all manner of disposition as "human resources" much as things present at hand (e.g., trees) and things ready to hand (e.g., tools) are appropriated to human use (Heidegger, 1977). The task before us, Heidegger urged, is to engage not the question, "What ought we to do?" but rather to engage the more fundamental question, "How must we think?" Unless this question is engaged, I submit, we may indeed find *homo sapiens sapiens* as we know him displaced and replaced by

human version 2.0. The call to engage this question goes out not only to contemporary philosophers and ethicists but also to scientists and neuroengineers whose empirical projects posit domains of inquiry without attention to foundational problems.

References

American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, 1994.

Arendt H. Eichmann in Jerusalem: A Report on the Banality of Evil (New York: Penguin Classics, 2006; Viking Press, 1963).

Bernstein R. From Radical Evil to the Banality of Evil: From Superfluousness to Thoughtlessness. in Hannah Arendt and the Jewish Question (Cambridge: The MIT Press, 1996), pp. 137-153

de Garis H. Building Gods or Building Our Potential Exterminators?. Published originally on 26 February 2001, available in electronic version online at: http://www.kurzweilai.net/meme/frame.html?main=/articles/art0131.html; accessed on 07 January 2008.

de Garis H. Cosmism: Nano Electronics and 21st Century Global Ideological Warfare. 1996; available at: http://psoup.math.wisc.edu/extras/deGaris/cosmism.html; accessed on 07 January 2008).

de Garis H. The 21st Century Artilect Moral Dilemmas Concerning the Ultra Intelligent Machine. Revue Internationale de Philosophie 1990;44(172):131-138.

de Garis H. The Artilect War: Cosmists vs. Terrans (Palm Springs CA: ETC Publications, 2005). (Machines Like Us Interviews: Hugo de Garis, 09/03/07; available online at http://www.machineslikeus.com/cms/interview-hugo-de-garis.html; accessed on 07 January 2008).

Greene J. The Secret of Kant's Soul. http://www.wjh.harvard.edu/~jgreene/; pp. 2-3; accessed on 12 January 2008.

Haidt J and Graham J. When Morality Opposes Justice: Conservatives Have Moral Intuitions That Liberals May Not Recognize. Social Justice Research 2007;20: 98-116.

Haidt J and Joseph C. The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules. In P. Carruthers, S. Laurence, and S. Stich, eds., The Innate Mind, Vol. 3), Chap 19, in press).

Haidt J and Joseph J. Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues. Daedalus, Fall 2004, pp. 56-66.

Haidt J. The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. Psychological Review 2001;108; 814-834.

Haidt J. The New Synthesis in Moral Psychology. Science 2007;316:998-1002.

Heidegger M. On Time and Being, trans. J. Stambaugh (New York: Harper & Row, 1972).

Heidegger M. The Question Concerning Technology, trans. W. Lovitt (San Francisco: Harper & Row, 1977).

MacIntyre A. Dependent Rational Animals: Why Human Beings Need the Virtues (Peru Illinois: Open Court Press/Carus Publishing Co., 2001).

MacIntyre A. Whose Justice? Which Rationality (Notre Dame: University of Notre Dame Press, 1989).

Morelli EM. Ressentiment and Rationality. 20th World Congress of Philosophy, Philosophical Anthropology Contributed Papers Section, 1999, available online at: http://www.bu.edu/wcp/Papers/Anth/AnthMore.htm).

Pinker S. The Moral Instinct. The New York Times Magazine, 13 January 2008; available online at http://www.nytimes.com/2008/01/13; accessed 13 January 2008.