ARTICLE

.1

Risk Mysterianism and Cognitive Boosters

Philippe Verdoux
Institute for Ethics and Emerging Technologies
USA

Abstract

In this paper, I argue that cognitive enhancements could increase our ability to analyze and control risks, most importantly existential risks. Towards this end, I identify two significant constraints on our ability to evaluate/neutralize such risks, one of which concerns the problem's type and the other its size (or complexity). I provide two examples of these constraints, and examine how enhancements could help overcome each.

Keywords: existential risks, future studies, secular eschatology, cognitive closure

Section 1

> *However forward-looking we may all pretend to be, humanity is far more interested in its past than the future.*
>
> *– John Mortimer*[1]

The philosopher George Santayana (1920) famously stated that "those who cannot remember the past are condemned to repeat it." This suggests that history exhibits a *contingent* cyclicality, and its moral is that we must learn from our mistakes if "progress" is to be made (Santayana, 1920, pp.284-285). But there are some sorts of mistakes from which we cannot possibly learn; mistakes that preclude retrospective analysis. Consider the difference between (what may be called) "historically unprecedented" and "historically singular" kinds of unrealized risks: the former are general types that can have more than one token, although no such instances have yet occurred,[2] while the latter are types that can only be instantiated once. A local act of nuclear terrorism, for example, would be a profound tragedy (as was the bombing of Hiroshima and Nagasaki), but humanity would survive it. As of April 2010, just after President Obama's promising two-day Nuclear Security Summit with 47 nations, no such act has yet taken place. Until this counterfactual becomes factual, then, it falls within the "unprecedented" category.

In contrast, historically singular risks are not just unprecedented but by stipulation non-repeatable.[3] The subclass of such occurrences that will concern us here are what Nick Bostrom calls existential risks, or "[risks] where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential" (Bostrom, 2002). Ontologically, while some existential risks derive from nature – e.g., an asteroid such as that imputed to have caused the Cretaceous-Tertiary extinction event – virtually *all* existential risks today are "technogenic" in nature. That is, they arise from the negative externalities and dual-usabilities of neoteric technologies, especially those associated with the inchoate genetics, nanotechnology and robotics (GNR) revolution (Forge, 2010; Kurzweil, 2005; Miller & Selgelid, 2007). In terms of disciplinary ownership, both the unprecedented and singular types of risks, by virtue of being unactualized possibilities, fall within the field of Future Studies; but only the latter has its home in the subfield of "Secular Eschatology."[4]

In a recent publication on risks, Bostrom and co-author Milan Cirkovic analyze existential risks as being "transgenerational" in scope and "terminal" in intensity (Bostrom & Cirkovic, 2008; Figure 1): whereas aging killed every individual alive during, say, the adoption of the US Declaration of Independence on July 4, 1776, the human race nonetheless endured. Aging is thus "global" rather than transgenerational in scope. An apocalyptic scenario involving ecophagic nanobots with the capacity for self-replication, in contrast, would result in the permanent end of our *species* (as well as the biosphere; see Phoenix & Treder, 2008). On this analysis, then, both scope and intensity fall under the concept of *consequences*, which figures in the standard definition of a risk as "the probability of an event occurring multiplied by its consequences."

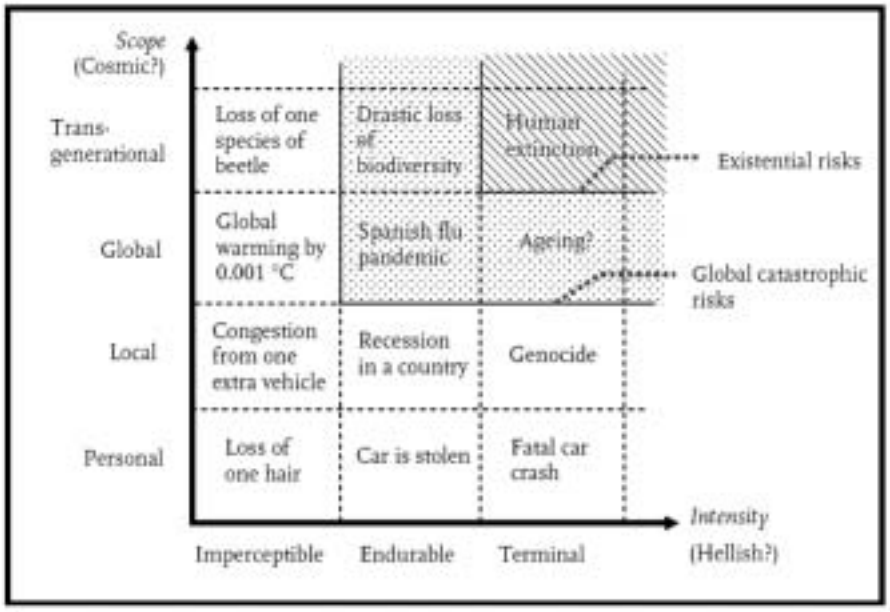

Figure 1. Typology of risks. (Bostrom & Cirkovic, 2008)

What is particularly alarming about existential risks is their rate of growth along two axes: number and probability (Verdoux, 2009). With respect to the former, Bostrom counts a total of 23 categories of existential risk kinds engendered by, for instance, nanotechnology, advanced AI systems, biotechnologies like genetic engineering and nuclear weapons. Furthermore, according to our best estimates, the probabilistic likelihood of an existential catastrophe being actualized in the 21st Century ranges from >25% (Bostrom, 2002) to >50% (Rees, 2003).

Note that these are not estimates that any one of the myriad "sub-existential" catastrophes will occur, such as a SARS-like global pandemic or a local act of nuclear terrorism. In fact, these "less worrisome" risks have a far greater likelihood of occurring. A 2008 report by the US's Commission on the Prevention of WMD Proliferation and Terrorism, for example, states that "it is more likely than not that a weapon of mass destruction will be used in a terrorist attack somewhere in the world by the end of 2013." And Harvard professor Graham Allison has estimated a >50% probability of a nuclear bomb exploding in a US city between 2004 and 2014.[5] Rather, the probability estimates given by Bostrom and Rees (as well as John Leslie and Richard Posner, to name a few others[6]) pertain to the grim possibility of complete (self-)annihilation. This sudden growth spurt of existential risks has, in fact, led the present author (2009) to speculate about the possibility of an "existential risk singularity" (ERS). An ERS would thus be (roughly speaking) a "hypothetical" future point at which the introduction of new existential risk types would vastly exceed our current capacities of comprehension.[7]

An important reason for such pessimism (although see Section 4) concerns the availability of both the technical and epistemic resources necessary for actualizing an existential catastrophe. Late modernity is, conspicuously, the Age of Information; unfortunately, some of this information can be used to harm humanity.[8] For example, a nefarious "biohacker" could easily access the specialized knowledge and instrumentation needed to independently synthesize a designer pathogen, like the Polio virus (see Pollack, 2002), in his or her basement laboratory. Thus, reaping massive destruction no longer requires a multi-person team of collaborating scientists, as the manufacture of an improvised nuclear device (IND) still does.[9] Instead, the human race could now be extirpated by a single psychopath – or group, such as the Gaia Liberation Front[10] – working in complete isolation. Imagine if Timothy McVeigh, who perpetrated the infamous Oklahoma City bombing, had been a competent microbiologist. Such is the next generation of technoscientific terrorists.

But intentional terror is not the only cause for apocalyptic anxiety: there is also the ominous possibility of unintended error, as occurred in 2007 when a military plane unknowingly transported six nuclear warheads on cruise missiles from bases in North Dakota to Louisiana (Star, 2007). Similar stories are easy to find: for example, on January 21st, 1968, a B-52 bomber carrying nuclear weapons crashed before the crew could radio SAC HQ (Strategic Air Command Headquarters). The anti-proliferationist Alan Phillips writes that "at that time, the 'one point safe' condition of the nuclear weapons could not be guaranteed, and it is believed that a nuclear explosion could have resulted from accidental detonation of the high explosive trigger." Thus, "had there been a nuclear detonation [the resultant explosion] would have given an indica-

tion consistent with a successful nuclear attack" (Phillips, 1998; see also Sagan, 1993). An all-out nuclear war might have ensued. Technological and human error is, therefore, just as grim as the grim threat of terror.[11]

In sum, while history may indeed be marked by thematic repetition (war, rebellion, tyranny, oppression, and so on), the present period of human civilization is historically unique in its invention of a brand new and dismal theme: *technogenic existential risks*. The question, then, is what to do about such risks: How can we prevent the further creation of new risk-types? and How can we ensure that risks already existing will not be actualized, through either error or terror? In my view, the special problem posed by existential risks should occupy a position atop our collective list of "the highest priorities." Future Studies in particular ought to focus more academic attention on secular eschatology issues (i.e., existential risks[12]), and more money ought to be funneled into supporting such research. As Bostrom has lamented, "it is sad that humanity as a whole has not invested even a few million dollars to improve its thinking about how it may best ensure its own survival" (Bostrom, 2006). I would, indeed, urge that there are not only compelling ethical reasons, given the universal imperative to reduce human suffering whenever and wherever possible, but robust practical reasons as well for funding/pursuing such research. After all, what good is composing a musical masterpiece if there's no one around to hear it? and What good is discovering whether the Higgs boson particle (in the case of the LHC) exists if there's no one around to know about it?

The average duration of mammalian species is roughly 2.2 million years (Avise, Walker & Johns, 1998). It appears that humanity may bring this average down through a kind of "colony collapse disorder" of our own.

Section 2

The human being: a genius with an idiot's brain.

– Anonymous

In a recent interview, a question "on whether we are capable as a species of tackling climate change" was posed to the Gaia theorist James Lovelock. In response, Lovelock opined: "I don't think we're yet evolved to the point where we're clever enough to handle [as] complex a situation as climate change" (Hickman, 2010). Implicit in this response is the notion that *if only we were smarter, then we'd be able to solve the problems confronting us* (Verdoux, 2010a). This is, in fact, a common and quite intuitive idea. Bostrom and Sandberg, for example, assert in an argument for the development of "cognitive enhancement technologies" that "society faces many pressing problems which would be more readily solved if its members were smarter, wiser, or more creative" (Bostrom & Sandberg, 2009). Similarly, the transhumanist Mark Walker claims that "the best candidates amongst us to lead civilization through such perilous times are the brightest and most virtuous: posthumans" (Walker, 2009). I myself have defended a similar thesis (Verdoux, 2009). But is this a tenable position? Would increasing our core cognitive capacities (memory, attention, perception, etc.) with things like nootropics, tissue grafts and neural implants (Walker, 2008) better

enable us to understand and control existential risks? Would the world become safer or even more dangerous?

On the one hand, just as some philosophers have maintained that there exist questions that we can ask but cannot *in principle* answer (McGinn, 1989; see below),[13] one might conjecture that there exist problems – social, economic, political, technological in nature – that we can create but simply cannot solve. Maybe we are clever enough to create a mess, but not to clean it up. It apparently follows, then, that a sufficiently strong cognitive booster could provide us with the mental resources needed to ameliorate our predicament – that is, to reverse global warming, adopt a sustainable *modus vivendi*, clean up the Great Pacific Garbage Patch and exorcize the specter of existential catastrophe that haunts the present millennium.

One could argue in response, though, that anthropological considerations suggest an upgrade of our mental software would actually *exacerbate* our plight, despite our pre-theoretic intuitions. Why? Because human history, beginning ~2.5 million years ago with *Homo habilis*, evinces an apparently strong correlation between increases in human intelligence (i.e., through the process of "encephalization") and increases in our actual or potential capacity for (self-)destruction (see Figure 2; Verdoux, 2010a). The extinction of the Pleistocene megafauna, for example, was largely catalyzed by humans, according to the overkill hypothesis. And this "intercontinental holocaust" occurred during a time of significant cognitive evolution, as both anatomical and archaeological studies show (see Ambrose, 2001).[14] The permanent erasure of the megafauna, as well as Neanderthals[15] and other such creatures, then bled into what biologists today call the "Holocene extinction event," which commenced with the Neolithic revolution roughly 10,000 years ago – another great innovation of human technology.

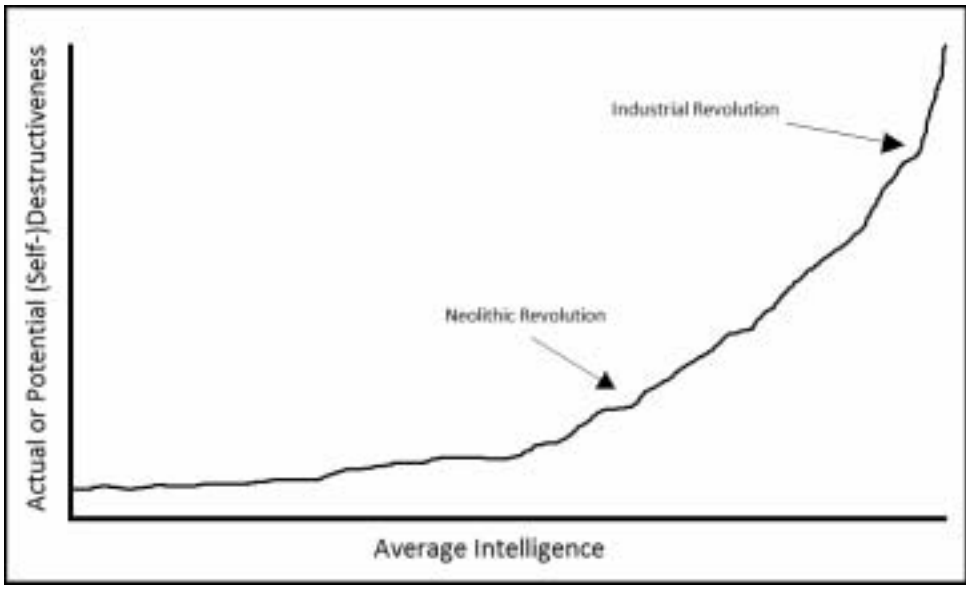

Figure 2. Diagram showing the historical relation between average intelligence and our species' capacity for (self-)destruction.

At present, in the midst of the "Anthropocene" (beginning with the European Industrial Revolution), scientists estimate that one in three species is at risk of extinction, about one fourth of all bird species on Earth have gone extinct, roughly half of the world's primate species are endangered and climatologists are warning of "irreversible' climate shifts because worst-case scenarios warned of two years ago are being realized" (CNN, 2009).[16] Furthermore, there is the growing malignancy of existential risks arising from neoteric technology discussed above. The punch line is that simultaneous to this unequivocal rise in our actual and potential (self-)destructiveness, psychologists have observed appreciable decadal increases in average IQ among citizens of the "developed" world. This is called the Flynn effect (Flynn, 2007). Thus, we are currently (i) impacting the planetary spaceship on which we live more than ever, and (ii) more intelligent than any Earth-originating organism has ever been. Maybe the *best* way out of the labyrinth we've created, therefore, would be to emulate the humble sea-squirt and "eat our own brains."[17]

Humans are not only "super-predators" at the top of the food chain but destroyers of the systems upon which we depend for survival; a more descriptively accurate binomen might be: *Homo annihilatus*. Thus, if we went extinct, as members of the Voluntary Human Extinction Movement (VHEMT[18]) wish, the effect on the global ecosystem would ultimately be quite positive. Although I concede this point, I nonetheless adopt a position similar to that defended by Bostrom and Sandberg, and Walker above, given my particular axiological orientation (which is not anthropocentric but nonetheless positively values the existence of *Homo sapiens*). We are, it seems, going through an awkward phase in the adolescence of our species; as David Brin notes, "we now have a new, tentative value system that's arisen in the most recent generation of the Modern West, wherein some initial signs of *self-restraint and satiability* have started to appear" (Brin, 2010). What we need, then, is to foster these incipient tendencies to maturity – that is, to become "more evolved," as Lovelock suggests – and the creation of cognitive enhancements provides the only feasible way of accomplishing this.[19]

Consider the following points. First, the groups most likely to perpetrate an act of terrorism – including one with existential consequences – are, as it were, "apocalyptic" and "politico-religious" in nature (Ferguson & Potter, 2004, pp.18-19). Both are motivated by highly irrational beliefs about how the world *ought to* be, and this makes for a lethal combination when advanced technologies enter the mix. As Sam Harris has stated: "I am terrified of what seems to me to be a bottleneck that civilization is passing through. On the one hand we have 21st-century disruptive technology proliferating, and on the other we have first-century superstition" (Harris, 2004; NW, 2007). A number of recent studies, though, report a statistically significant link between intelligence and atheism (Kanazawa, 2010; Lynn, Harvey & Nyborg, 2009[20]). Thus, if one accepts the conclusions of these studies, it follows that by increasing intelligence via cognition-enhancing means (as simple as a better education and as sophisticated as mind-uploading), the number of individuals espousing religious dogma would consequently decline, thereby reducing the probabilistic threat of terrorism (see Verdoux, 2010d for more).

Furthermore, the study of existential risks itself suffers from a number of psychological biases that can sway one's opinion about a given risk. As mentioned above, "historically singular" risks are all *futurological* in nature, which means that probability estimates like those given by Bostrom and Sir Martin Rees (in Section 1) are necessarily subjective in nature. This leaves room for cognitive biases to creep in and insidiously influence our judgments. Bostrom, for example, discusses "observation selection effects," which may lead one to underestimate the likelihood of a catastrophe (Tegmark & Bostrom, 2005), as well as (what he calls) the "good-story bias." The latter is, Bostrom explains, a tendency to think of futuristic scenarios with exciting narratives as being more likely to occur than those with boring ones (Bostrom, 2002). There is also the "numerator bias," which leads one to judge that saving 7/100 lives is better than 1/10, even though the former is a lower percentage (Slovic, Finucane, Peters & MacGregor, 2009). For the present purposes, given the oceanic literature on the subject, it suffices to note that cognitive enhancements could help neutralize such biases, thus allowing the sober eschatologist to better estimate the probability of a secular Armageddon.

Section 3

> *It is no good to try to stop knowledge from going forward. Ignorance is never better than knowledge.*
>
> *– Enrico Fermi*[21]

In the remainder of this essay, I would like to focus on two specific constraints on our ability to analyze risks. These are plausibly soluble, I believe, by cognitive enhancement technologies; and thus, given the urgency of our existential situation today, we have good reason for developing them. In other words, if Lovelock is correct and we are not sufficiently clever at present to solve various eschatology-sized macro-problems like global warming, and if cognitive enhancements have a high probability of enabling us to be sufficiently clever, then it behooves us to immediately direct as many resources – including money and scientists – toward the development of cognition-boosters. But this may mean postponing "blue skies" projects, such as the LHC, despite the high intellectual value they may possess (Verdoux, 2010b). The issue, therefore, is ultimately one of prioritization, all things considered. As far as I can see, cognitive enhancements may be our only hope for surviving the hostile selective environment that we, in our foolish youth, have collectively constructed.[22] To begin, then:

Science, a human endeavor, has insuperable limits (Barrow 1998). The Heisenberg Uncertainty Principle, for example, states that "the product of the uncertainty in position and the uncertainty in momentum of a particle *can be no smaller than* Plank's constant divided by 4π" (Ebbing & Gammon, 2005[23]). Analogously, Godel's incompleteness theorems in mathematical logic reveal principled limitations inherent in axiomatic systems. What concerns us at present, though, are two specific "sources of ignorance" (Verdoux, 2010c), one of which the "New Mysterian" philosopher Colin McGinn calls "cognitive closure" and the other of which is closely related

but non-identical. As McGinn puts it, "a type of mind *M* is cognitive closed with respect to a property *P* (or theory *T*) if and only if the concept-forming procedures at *M*'s disposal cannot extend to a grasp of *P* (or an understanding of *T*)" (McGinn, 1989, p.350). In other words, there may be theories requiring one to possess certain concepts that are permanently beyond our epistemic reach (but maybe not the reach of some other species, say, *Extra terrestrialus or Gedanken experimentus*). Call this the problem of type.

In contrast, some problems are abstruse not because they include concepts too difficult for the human mind to grasp, but because they involve too many component parts for the human mind to keep in order. For example, a leaf falling from a tree branch in autumn follows deterministic laws. But a physicist trying to describe the leaf's movement would find it impossible to do so. Why? Simply because the process is far too complicated to map out in detail. Call this the problem of size (or complexity).[24] While the type problem has and always will beleaguer *Homo sapiens*, the latter has become increasingly consternating as our world continues to complexify exponentially (see Modis, 2002); such complexification is manifest in, for example, the phenomenon of specialization.[25] In fact, one recent study concluded that "the fastest growing entity today is information" (Kelly, 2008). By all accounts, then, information-overload is a real worry,[26] especially when certain information may decide whether the present epoch turns out to be – as Bertrand Russell once put it – humanity's "prologue or epilogue."[27]

Now, both the type and size problems are, as stated, relevant to risk analysis and control. Their resolution would thus increase our ability to detect and understand risks, and then to devise efficacious strategies for neutralizing them. Consider, for example, the case of global warming. By all accounts, global warming is a messy problem of extraordinary complexity. A rudimentary understanding of its etiology, for example, requires one to know *something* about physics, chemistry, biology, ecology, meteorology and climatology, not to mention the economic, geopolitical, and so on, variables that also figure in the causal equation.[28] Without going into great depth on the subject (in part because of its size), consider a few of the proposed "geoengineering" solutions. These solutions, in themselves, pose significant epistemic hurdles for scientists and policy makers to jump. For example, one suggestion is to "fertilize" the oceans with iron to foment the growth of algal blooms (the result of rapid increases in phytoplankton biomass) (Cascio, 2010). This would decrease atmospheric CO_2 and, consequently, increase the amount of infrared radiation (heat) that could escape back into space. On the down side, though, this strategy might exacerbate ocean eutrophication; in fact, there are at present over 400 "dead zones" worldwide, "affecting a total area of more than 245,000 square kilometers" (Diaz & Rosenberg, 2008). Yet another problem of global catastrophic proportions.

Another proposal is to inject the stratosphere with sulfur dioxide or aluminum oxide – or even some "designer self-levitating aerosols that might be engineered to migrate to particular regions (e.g. over the arctic) or to rise above the stratospher[e] (so as not to interfere in stratospheric chemistry)" (Ricke, Morgan & Apt, 2008) – in an effort to enhance planetary albedo. The idea here is that even though more heat energy would be still trapped by the greenhouse effect (as a result of continuing fossil

fuel combustion), less light would reach earth to be reradiated in the first place. Others researchers, including the Working Group III of the Intergovernmental Panel on Climate Change (IPCC), have discussed the alternative of developing carbon capture and storage (CCS) devices at "large point sources," in an effort to decrease the quantity of greenhouse gases (GHGs) released into the atmosphere. Such apparatuses would involve "the separation of CO_2 from industrial and energy-related sources, transport to a storage location and long-term isolation from the atmosphere" (WG, 2005, p.3).

The best solution would undoubtedly entail the creation of new, cutting-edge technologies to provide alternatives to fossil fuel combustion and remove anthropogenic GHGs already in the atmosphere, while at the same time maintaining an acceptable level of economic prosperity.[29] But can we devise such technologies fast enough – that is, before something *really* bad happens? At the extreme, climate changes brought about by us self-described "wise apes" may reach a "climate threshold," thus initiating a catastrophic "runaway greenhouse effect," as may have occurred on Venus (Philander, 2008, p.225). Clearly, having more bright minds and more minds that are brighter would facilitate progress towards the end of "mitigation and adaptation" (IPCC, 2007). Cognitive enhancements promise to augment exactly the sort of mental capacities, such as attention, memory and even creativity (Orca, 2009), needed to devise good solutions for such sizable problems.[30]

Now, consider the LHC.[31] According to the LHC Safety Assessment Group (LSAG) report, there is virtually zero chance that any of CERN's four experiments at the LHC will create a microscopic black hole, strangelet or vacuum bubble (LSAG, 2008). To begin, it is always judicious to remind oneself of the intrinsic fallibility of science, as well as the historical truism that paradigm shifts sometimes effectuate radical changes in our most basic orientation towards reality. (Recall that Egas Moniz once won science's most prestigious honor, the Nobel Prize, for the invention of the lobotomy, or that phlogiston was universally accepted by chemists until the late eighteenth century.) As Max Planck once noted, "a new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it" (see Kuhn, 1962).[32] Thus, for all we know right now, the next generation of physicists, working within a modified framework of more advanced theory, *will* regard the LHC's risks as significant. This should, I believe, give insatiable "epistemivores" (including myself) more pause than it typically does.

Second, the LHC's experiments are located at the furthest edge of our actual – and maybe *possible* – understanding of the cosmos, given the limited "concept-forming procedures" at our disposal. In McGinn's words:

> *One of the areas in which the possibility of cognitive closure looks most real is theoretical physics – quantum theory and the origin of the universe being the standard examples. The more advanced a theory becomes the more likely it is to approach the limits of what we can know. (McGinn, 1991, p.88)*

This idea becomes especially vivid when one hears LHC physicists, such as Brian Cox, describe the project as "certainly, by far, the biggest jump into the unknown" that science has *ever* attempted (BBC, 2007). Central to the issue are the conceptual diffi-

culties posed by advanced physical theories like string theory, which posits the existence of 11 dimensions in which one-dimensional "strings" vibrate at different frequencies (each vibrational spectrum corresponding to a different particle, such as the graviton[33]). As Rick Groleau notes in a NOVA article on the subject, "even physicists who spend all day thinking about extra spatial dimensions have a hard time describing what they might look like or how we apparently feeble-minded humans might approach an understanding of them." Scientists are, after all, only human. Groleau then concludes: "That's always been the case, and perhaps always will be."

But cognitive closure is a *species-relative* phenomenon. Thus, if a more "advanced" species – call them "posthumans," for lack of a better term – were to be created through the integration of biology and (cognitive enhancement) technology,[34] then surely the venerable boundary between the knowable and the unknowable *for humans* could be redefined *for them* (see Figure 3). Our posthuman progeny may, as a result, understand concepts that are as unintelligible to us as the notion of "social justice" is to the chirping cricket.[35]

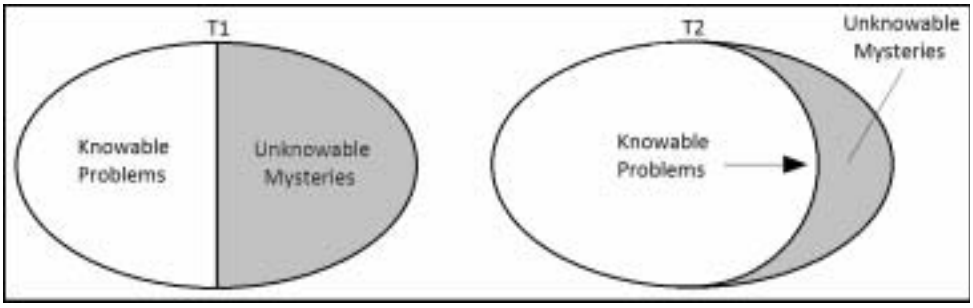

Figure 3. Cognitive enhancement technologies offer the possibility of redefining, from *T1* to *T2*, the boundary between "problems" and "mysteries." (Not drawn to scale.[36])

With respect to risks, then, it is not unreasonable to conjecture that certain risks might only be epistemically "visible" from the vantage point of some theory *T* – a theory of which we are not only unaware at present, but with respect to which we are permanently "closed," in the cognitive sense. By technologically redefining the line between what Noam Chomsky has independently termed "problems" and "mysteries," therefore, our ability to provide more accurate and comprehensive assessments of the risks involved in LHC-like experiments would be improved.[37] And if such improvement were to occur, maybe the LSAG report "Version II" would end up being quite different. We just can't know as long as we've got the brains we've got.

Section 4

It seems to me there is all the difference in the world between those who profess to believe in progress and those who would work to achieve it.

– Dale Carrico

The thesis here defended is that: (i) the study of existential risks ought to be a top priority for Future Studies and other related fields (such as cosmology, technology studies, environmental science, and so on); and (ii) given their potential for increasing our capacities of risk analysis and control, in combination with the apparent urgency of our existential situation, R&D of cognitive enhancements ought to – ideally – take precedence over any project that does not have a *high probability* of yielding results that could help mitigate the threat of existential disaster. There are, once again, compelling practical and ethical reasons for adopting this position.

I would like to end by distinguishing between (what I call) "theoretical pessimism" and "practical optimism" – a distinction upon which the present thesis rests. The former is what results from a careful examination of the empirical facts and the best available prognostications of the future. It turns out, I have suggested, that there are fairly cogent reasons for thinking that the Great Filter is temporally in front of rather than behind us (see Hanson, 1998). Indeed, according to some interpretations of the Drake equation,[38] a high likelihood should be assigned to the possibility that intelligent lifeforms populate the universe, and thus that we are not living on a "rare earth" (Ward & Brownlee, 2000). Yet the verifiable evidence for the existence of aliens is virtually zero. Maybe this is because, some have speculated, most exoplanetary beings of high intelligence reach a "post-industrial" point in their civilization's history at which the collective is just smart enough to actualize an existential risk, but not to effectively prevent it. Call this general hypothesis "Risk Mysterianism," i.e., "it's a *mystery* how to obviate the disasters we've created."

That being said, it is precisely the *haunting reality* of complete humanicide that motivates the second component of the above distinction: practical optimism. This proactive (or "proactionary") attitude rejects the fatalistic notion "that we are powerless to do anything other than what we actually do" (Rice, 2006),[39] and it embraces the possibilities of self-improvement through "enhancive" means. Again, the most quotidian might involve merely securing a decent education, getting regular exercise or maintaining good "neural hygiene," such as by avoiding the positive and negative causes of Environmentally Mediated Intellectual Decline (EMID) (Williams, 1999). At the extreme limits of human cognition, I have identified two sources of ignorance – the problems of type and size – that must be targeted by more radical cognitive enhancements. Neutralizing these ignorance-sources would help us solve a range of problems, from global warming (a size problem) to potentially risky physics experiments (a type problem) – and everything in-between.

There is, in conclusion, much to be enthusiastic about if only we sublimate our theoretical pessimism into practical optimism.

Correspondence

Philippe Verdoux
Institute for Ethics and Emerging Technologies
New York, USA
711 W. Carver St. / Durham, NC. 27704
Tel: 240.422.1123
E-mail: philippe.verdoux@gmail.com

Notes

1. Quoted in Leslie (1996, vii).
2. There is, of course, a crucial deictic element here. As I mention below, once such a risk is realized, it no longer falls within the category of "unrealized risks."
3. There could, of course, be non-repeatable risks that are humanly endurable. The extinction of the Mauritius dodo, for example, was a one-time event (assuming that future genetic engineers are unable to resurrect the species); and while this event had consequences for the local island ecosystem, such consequences could not have endangered the continued existence of *Homo sapiens*.
4. Within Secular Eschatology is an already-established subfield of cosmology called "Physical Eschatology," which studies the entropic end of the universe a "heat death" whimper.
5. I take this figure from a lecture found at the following link: URL = <http://www.youtube.com/watch?v=qUYrdCmfSSE>.
6. See Posner 2004 and Leslie 1996.
7. See Kurzweil 2005 for more on The Singularity, a technological phenomenon.
8. Bostrom has posted a draft of a paper entitled "Information Hazards: A Typology of Potential Harms from Knowledge" (2009b), in which he discusses the risks associated with certain kinds of information in detail.
9. The Manhattan Project was, of course, the first Big Science project; and CERN's Large Hadron Collider (LHC) now in operation boasts of being the biggest Big Science project in human history. Scientific collaboration over time seems to follow something like "Cope's rule."
10. In 1995, this group "recommended using biological weapons to extinguish humanity. It's manifesto stated: 'we can ensure Gaia's survival only through the extinction of the Humans as a species... we now have the specific technology for doing the job'" (Matheny, 2007, 1337, footnote 5).
11. Accidents are, of course, a ubiquitous part of life. In fact, accidents are the fifth leading cause of death in the US (Ropeik & Gray, 2002).
12. Or more millenialist-utopian possibilities like The Singularity (Hughes, 2008).
13. Nicholas Rescher makes a nice distinction between "merely unanswerable questions and actually unaskable ones" (Rescher, 1999, p.10). According to McGinn, conscious experience – the "what it's like to be" something – constitutes a question that we can pose but lack the requisite mental machinery to answer. But one need not accept this *particular* claim about us humans being cognitively closed to the mind-body problem for one to accept cognitive closure to *some* domain of questions as a real feature of our biological predicament.
14. I borrow the term "intercontinental holocaust" from Weisman 2005.
15. Although, since writing this paper, a study was published that strongly suggests interbreeding occurred between Neanderthals and non-Africans. See Keim 2010.
16. Since writing this paper, the Convention on Biological Diversity released their third *Global Biodiversity Outlook (GBO-3)* report. Among an alarming list of environmental tragedies, *GBO-3* states that "the abundance of vertebrate species, based on assessed populations, fell by nearly one-third on average between 1970 and 2006, and continues

to fall globally, with especially severe declines in the tropics and among freshwater species," as well as that "crop and livestock genetic diversity continues to decline in agricultural systems. For example, more than sixty breeds of livestock are reported to have become extinct since 2000" (UNEP, 2010).

17. Obviously, correlation does not imply causation. Nonetheless, I believe that closer examine of the causal relations between increasingly intelligent humans and our impact on the globe reveals an unequivocal causal link, at least historically.
18. This is pronounced [**vee**-*uh*-m*uh* nt].
19. While Darwinian natural selection has increased our intelligence in the past, there is currently no selective pressure in our environment for more intelligence. In fact, studies show that those with lower IQs tend to have more children than those with higher IQs.
20. I should add that while I find any remarks in these papers, or by these authors in other publications, concerning a racial basis for intellectual ability completely unfounded, methodologically flawed and morally offensive, I cannot detect any problems with the within-culture correlations between IQ measurements and atheistic beliefs. These seem robust, as far as I can tell, despite whatever additional absurdities the authors may espouse. I am attempting to avoid the "genetic fallacy" – no pun intended.
21. Quoted in Webb (2002, p.8).
22. Although space colonization may also reduce the probability of existential disaster, since "a species' survivability is closely related to the extent of its range" (Matheny, 2007, p.1337). I should also make explicit that the present paper brackets a number of *highly important* ethical issues relating to cognitive enhancements, which are almost all dual-use in nature. The present paper merely attempts to emphasize the very real positive benefits of such technologies; a subsequent paper will deal with the potential downsides. (See also Jonathan Moreno's consternating 2006 book entitled *Mind Wars*.)
23. Emphasis modified.
24. McGinn makes this distinction in the context of subjective experience. McGinn writes that "it is not the *size* of the problem but its type that makes the mind-body problem so hard for us" (McGinn, 1989, p.364).
25. This is exactly what makes John Hodgman's (2005) encyclopedic book so risible. The full title is: *An Almanac of Complete World Knowledge Compiled with Instructive Annotation and Arranged in Useful Order by myself, John Hodgman, a Professional Writer, in* ***The Areas of My Expertise****, which Include: Matters Historical, Matters Literary, Matters Cryptozoological, Hobo Matters, Food, Drink & Cheese (a Kind of Food), Squirrels & Lobsters & Eels, Haircuts, Utopia, What Will Happen in the Future, and Most Other Subjects.*
26. As I have elsewhere put it, *everyone today knows almost nothing about most things* (Verdoux, 2010c).
27. Irvine 1998.
28. This is partly why so many non-experts in the US, even those who value intellectual honesty, remain skeptical of global warming. For example, the 2009-2010 winter was the snowiest on record on the US East Coast. This led many, in particular the more radical "global warming fideists" associated with Fox News (such as Steve Milloy), to declare that Al Gore's "hysterical" theory of global warming was, obviously, false. Yet it turns out that January 2010 was, globally speaking, the warmest on record (Hudson,

2010). A failure to distinguish between global and local trends (which may not always be identical) is, for most people, merely a product of ignorance.

29. There is also the less plausible possibility of an "attitudinal" rather than "technological" fix. Downregulating our consumeristic appetites and the dominant "technological mood" that "enframes" the world as a vast resource waiting to be exploited (Heidegger, 1977) would clearly have a positive impact on our predicament. But this seems unlikely at present – we are not "evolved" enough for such disciplined self-control, even if hints of this are evident (see Brin, 2010).
30. There is good reason to be circumspect here. First of all, as mentioned in footnote nineteen, much of the research behind cognitive enhancements is dually usable (Moreno, 2006). And second, as Bostrom and Sandberg observe, "even therapeutic medicine, based on fairly good data from large clinical trials, is [very] hard to get right" (Bostrom & Sandberg, 2007, p.376). According to a recent study, in fact, iatrogenic death is the third leading cause of death in the US (Starfield, 2000). But one could turn these data around and argue that *this is precisely why more resources should be diverted toward enhancement research*. Surely if the "blind watchmaker" can engineer intelligent life, which it did, so too can the intelligent life it engineered.
31. While I used the LHC as an example of the type-problem, it also presents a number of extraordinarily sizable problems.
32. Or consider Clarke's first of three "laws of prediction": "When a distinguished but elderly scientists states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong."
33. See Weinstein 2005 for helpful discussion.
34. Although according to Bostrom (2008), posthumans could also be engendered by enhancements that are non-cognitive in nature, including ones that target our emotionality and healthspan.
35. An important analogue here pertains to the instrumentational embodiment of science (Ihde, 1991). That is, just as cognitive enhancements could provide epistemic access to realms of knowledge previously hidden to us, so too do the instruments of science "extend" our sense organs to provide perceptual access to domains of phenomena also previous hidden to us. As Thomas Kuhn (1962) suggests, technological innovation has often led scientific theorization.
36. This illustration assumes that the human mind is limited, but expandable, and that the universe is limited too (i.e., that there is something like a "final theory" of the cosmos). This seems to me to be the best account of our epistemic situation in the universe. But there are other options: one could, alternatively, hold that the human mind and the universe are both unlimited; that the human mind is limited but the universe is not; or that the universe is limited but the human mind is not (see Barrow, 1998, pp.72-82).
37. Chomsky 1975.
38. Frank Drake himself estimates that there are "about 10,000 communicating civilizations in the Milky Way alone" (Plaxco & Gross 2006, p.247).
39. One finds echoes of a fatalistic inevitability in singularitarianism. For example, Kurzweil proposes a linear 6-stage model of cosmic evolution that unfolds according to a transhistorical "law of accelerating returns." In my view, one *can* reasonably assert that the GNR (genetics, nanotechnology and robotics) revolution is inevitable, but only after

adding a number of strong *ceteris paribus* conditions, thus yielding the peculiar locution: "*ceteris paribus* inevitability." Obviously, our future may contain a secular apocalypse rather than a secular utopia, or neither (Bostrom, 2009a).

References

Ambrose, Stanley. (2001). Paleolithic technology and human evolution. *Science, 291*, 1748-1753.

Avise, John L, DeEtte Walker, & Glenn Johns. (1998). Speciation durations and pleistocene effects on vertebrate phylogeography. *Proceedings of the Royal Society of London Series B-Biological Sciences, 265*, 1707-1712.

Barrow, John. (1998). *Impossibility: The limits of science and the science of limits.* New York: Oxford University Press.

BBC. (2007). The six billion dollar experiment. Retrieved from http://www.bbc.co.uk/sn/tvradio/programmes/horizon/broadband/tx/universe/highlights/

Bostrom, Nick. (2002). Existential risks: Analyzing human extinction scenarios and related hazards. *Journal of Evolution and Technology, 9*.

Bostrom, Nick. (2006). Dinosaurs, dodos, humans? *Global Agenda*, January, 230-231.

Bostrom, Nick. (2008). Why I want to be posthuman when I grow up. In Bert Gordijn & Ruth Chadwick (Eds.), *Medical enhancement and posthumanity*. New York: Springer.

Bostrom, Nick. (2009a). The future of humanity. In Jan Kyrre Berg Olsen, Evan Selinger (Eds.), *New waves in philosophy of technology*. New York: Palgrave McMillan. Retrieved from http://www.nickbostrom.com/papers/future.pdf

Bostrom, Nick. (2009b). Information Hazards: A Typology of Potential Harms from Knowledge. Retrieved from http://www.nickbostrom.com/information-hazards.pdf

Bostrom, Nick. & Milan Cirkovic. (2008). Introduction. In Nick Bostrom & Milan Cirkovic (Eds.), *Global catastrophic risks*. Oxford: Oxford University Press. Retrieved from http://www. global-catastrophic-risks.com/docs/Chap01.pdf

Bostrom, Nick. & Anders Sandberg. (2007). The wisdom of nature: An evolutionary heuristic for human enhancement. In Julian Savulescu & Nick Bostrom (Eds.), *Enhancing humans*. Oxford: Oxford University Press.

Bostrom, Nick. & Anders Sandberg. (2009). Cognitive enhancement: Methods, ethics, regulatory challenges. *Science and Engineering Ethics, 15*(3), 311-341.

Brinn, David. (2010). Perspectives on SETI and Aliens. *Open Salon*. Retrieved March 1, 2010, from http://open.salon.com/blog/david_brin/2010/05/04/perspectives_on_seti_and_aliens

Cascio, James. (2010). A survival guide to geoengineering. *Momentum*. Retrieved March 10, 2010, from http://www. environment.umn.edu/momentum/current/feature3.html

Carrico, Dave. (2005). Progress as a natural force versus progress as the great work. *Institute for ethics and emerging technologies*. Retrieved January 15, 2007, from http://ieet.org/index.php/IEET/ more/130/

Chomsky, Noam. (1975). *Reflections on language*. New York: Pantheon.

CNN. (2009). World faces "irreversible" climate changes, researchers warn. *CNN*. Retrieved March 9, 2010, from http://edition.cnn.com/2009/WORLD/europe/03/12/irreversible.climate/index.html

Cohen, Mark N. (1989). *Health and the rise of civilization.* New Haven: Yale University Press.

Diaz, Robert, & Rutger Rosenberg. (2008). Spreading dead zones and consequences for marine ecosystems. *Science, 321*(5891), 926-929.

Ebbing, Darrell, & Steven Gammon. (2005). *General chemistry*. New York: Houghton Mifflin.

Ferguson, Craig, & William Potter. (2004). *Four faces of nuclear terrorism.* New York: Routledge.

Forge, John. (2010). A note on the definition of "dual use". *Science and Engineering Ethics, 16*(1), 111-118.

Flynn, James. (2007). *What is intelligence? Beyond the Flynn effect*. Cambridge, UK: Cambridge University Press.

Hanson, Robin. (1998). The great filter: Are we almost past Ii? Retrieved October 3, 2009, from http://hanson.gmu.edu/greatfilter.html

Harris, Sam. (2004). *The end of faith: Religion, terror, and the future of reason*. New York: W.W. Norton & Company.

Heidegger, Martin. (1977). *The question concerning technology and other essays*. New York: Harper Torch.

Hickman, Leo. (2010). James lovelock on the value of sceptics and why copenhagen was doomed. *The guardian*. Retrieved March 20, 2010, from http://www.guardian.co.uk/environment/blog/ 2010/mar/29/james-lovelock

Hodgman, John. (2005). *The areas of my expertise*. Boston, MA: Dutton Adult.

Hudson, Paul. (2010). January 2010 warmest on record. Yes really! *BBC Weather*. Retrieved March 8, 2010, from http://www.bbc.co.uk/blogs/paulhudson/2010/02/january-2010-warmest-on-record.shtml

Hughes, James. (2008). Millennial tendencies in responses to apocalyptic threats. In Nick Bostrom & Milan Cirkovic (Eds.), *Global catastrophic risks*. Oxford: Oxford University Press.

Ihde, Don. (1991). *Instrumental realism: The interface between philosophy of science and philosophy of technology*. Indiana: Indiana University Press.

IPCC. (2007). IPCC fourth assessment report: Climate change 2007. Retrieved March 20, 2010, from http://www.ipcc.ch/publications_and_data/publications_and_data_reports.htm#1

Irvine, Andrew. (1998). *Bertrand russell: Critical assessments*. New York: Routledge.

Kanazawa, Satoshi. (2010). Why liberals and atheists are more intelligent. *Social Psychology Quarterly, 73*(1), 33-57.

Keim, Brandon. (2010). Neanderthal genome shows most humans are cavemen. *Wired*. Retrieved June 3, 2010, from http://www.wired.com/wiredscience/2010/05/neanderthal-genome/

Kelly, Kevin. (2008). The expansion of ignorance. *The technium*. Retrieved September 24, 2009, from http://www.kk.org/ thetechnium/archives/2008/10/the_expansion_o.php

Kuhn, Thomas. (1962). *The structure of scientific revolutions*. Chicago, Il: University of Chicago Press.

Kurzweil, Ray. (2005). *The singularity is near: When humans transcend biology.* New York: Viking.

Leslie, John. (1996). *The end of the world: The science and ethics of human extinction*. London: Routledge.

LSAG. (2008). Review of the safety of LHC collisions, LHC Safety Assessment Group. Retrieved March 1, 2010, from http://lsag.web.cern.ch/lsag/LSAG-Report.pdf

Lynn, Richard, John Harvey & Helmuth Nyborg. (2009). Average intelligence predicts atheism rates across 137 nations. *Intelligence, 37*(1), 11-15.

Matheny, Jason. (2007). Reducing the risk of human extinction. *Risk Analysis, 27*(5), 1335-1344.

McGinn, Colin. (1998). Can we solve the mind-body problem? *Mind, 98*(391), 349-366.

McGinn, Colin. (1991). Consciousness and content. In R. Bogdan (Ed.), *Mind and common sense: Philosophical essays on commonsense psychology*. Cambridge, UK: Cambridge University Press.

Miller, Seumas, & Michael Selgelid. (2007). Ethical and phiosophical consideration of the dual-use dilemma in the biological science. *Science and Engineering Ethics, 13*(4), 523-580.

Moreno, Jonathan. (2006). *Mind wars: Basic research and national defense*. New York: Dana.

NW. (2007). The god debate. *NewsWeek*. Retrieved March 10, 2010, from http://www.newsweek.com/id/ 35784/page/1

Orca, Surfdaddy. (2009). Will we eventually upload our minds? *Humanity plus*. Retrieved January 2, 2010, from http://www.hplusmagazine.com/articles/neuro/will-we-eventually-upload-our-minds

Philander, S. George. (2008). *Encyclopedia of global warming and climate change*. New York: Sage.

Phillips, Alan. (1998). 20 mishaps that might have started accidental nuclear war. *Nuclear Age Peace Foundation*. Retrieved March 15, 2010, from http://www.wagingpeace.org/articles/1998/01/00_ phillips_20-mishaps.php

Phoenix, Chris, & Mike Treder. (2008). Nanotechnology as global catastrophic risk. In Nick Bostrom & Milan Cirkovic (Eds.), *Global catastrophic risks* (pp.481-503). Oxford, UK: Oxford University Press.

Plaxco, Kevin, & Michael Gross. (2006). *Astrobiology: A brief introduction*. Baltimore, MD: The Johns Hopkins University Press.

Pollack, Andrew. (2002). Scientists create a live polio virus. *The New York Times*. Retrieved March 19, 2010, from http://www.nytimes.com/2002/07/12/science/12POLI.html?pagewanted=1

Posner, Richard. (2004). *Catastrophe*. Oxford, UK: Oxford University Press.

Rees, Martin. (2003). *Our final hour: A scientist's warning: How terror, error, and environmental disaster threaten humankind's future in this century – On earth and beyond*. New York: Basic.

Rescher, Nicholas. (1999). *The limits of science*. Pittsburgh, PA: University of Pittsburgh Press.

Rice, Hugh. (2009). Fatalism. In Edward N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. Retrieved March 4, 2010, from http://plato.stanford.edu/archives/fall2009/entries/fatalism/

Ricke, Katharine, M. Granger Morgan & Jay Apt. (2008). Unilateral geoengineering. Retrieved February 10, 2010, from http://www.scribd.com/doc/13725689/Unilateral-Geoengineering-2009

Ropeik, David, & George Gray. (2002). *Risk: A practical guide for deciding what's really safe and what's really dangerous in the world around you*. Boston: Houghton Mifflin.

Sagan, Scott Douglas. (1993). *The limits of safety*. Princeton, NJ: Princeton University Press.

Santayana, George. (1920). *The life of reason: Or, the phases of human progress*. Retrieved February 5, 2010, from http://books.google.com/books/download/The_life_of_reason.pdf?id=KNvWAAAAMAAJ&output=pdf&sig=ACfU3U31-6EsGbmpw22qdNA4TzJiHAp_YQ&source=gbs_v2_summary_r&cad=0

Slovic, Paul, Melissa Finucane, Ellen Peters & Donald MacGregor. (2009). Risk as feeling: Some thoughts about affect, reason, risk and rationality. *The Ethics of Technological Risk*. London: Earthscan.

Starr, Barbara. (2007). Air force investigates mistaken transport of nuclear warheads. *CNN*. Retrieved http://www.nytimes.com/2002/07/12/science/12POLI.html?pagewanted=1

Starfield, Barbara. (2000). Is US health really the best in the world? *Journal of the American Medical Association, 284*(4), 483-485.

Tegmark, Max, & Nick Bostrom. (2005). How unlikely is a doomsday catastrophe? *Nature, 438*, 754.

Verdoux, Philippe. (2009). Transhumanism, progress and the future. *Journal of Evolution and Technology, 20*(2), 49-69.

Verdoux, Philippe. (2010a). If only we were smarter! *Institute for ethics and emerging technologies*. Retrieved March 20, 2010, from http://ieet.org/index.php/IEET/more/verdoux20100320/

Verdoux, Philippe. (2010b). Blue skies and existential risks. *Institute for ethics and emerging technologies*. Retrieved March 7, 2010, from http://ieet.org/index.php/IEET/more/verdoux20100407/

Verdoux, Philippe. (2010c). Technology and our epistemic situation. Under review.

Verdoux, Philippe. (2010d). Will All Posthumans Be Atheists? *Institute for ethics and emerging technologies*. Retrieved from http://ieet.org/index.php/IEET/more/verdoux20100601/

Verdoux, Philippe. (2010e). *Emerging technologies and the future of philosophy*. Forthcoming in *Metaphilosophy, 42*(5).

Walker, Mark. (2008). Cognitive enhancement and the identity objection. *Journal of Evolution and Technology, 18*(1), 108-115.

Walker, Mark. (2009). Ship of fools: Why transhumanism is the best bet to prevent the extinction of civilization. *The global spiral*. Retrieved March 8, 2010, from http://www.metanexus.net/ magazine/tabid/68/id/10682/Default.aspx

Ward. Peter, & Donald Brownless. (2000). *Rare earth: Why complex life is uncommon in the universe*. New York: Springer.

Webb, Stephen. (2002). *If the universe is teeming with aliens... where is everybody? Fifty solutions to Fermi's paradox and the problem of extraterrestrial life*. New York, NY: Springer.

Weinstein, Steven. (2008). Quantum gravity. In E. Zalta (Ed.), *The stanford encyclopedia of philosophy*. Retrieved June 18, 2010, from http://plato.stanford.edu/archives/fall2008/entries/quantum-gravity/

Weisman, Alan. (2005). Earth without people. *Discover magazine*. Retrieved March 23, 2010, from http://discovermagazine.com/2005/feb/earth-without-people/article_view?b_start:int=1&-C=

WG. (2005). Carbon dioxide capture and storage: Summary for policymakers. Retrieved March 28, 2010, from http://www.ipcc.ch/pdf/special-reports/srccs/srccs_summary-forpolicymakers.pdf

Williams, Christopher. (1999). *Termninus brain: The environmental threats to human intelligence*. New York: Cassell.

Wright, Robert. (2004). *A short history of progress.* Toronto, ON: House of Anansi.