Frank J. Tipler

Inevitable Existence and Inevitable Goodness of the Singularity

Abstract: *I show that the known fundamental laws of physics — quantum mechanics, general relativity, and the particle physics Standard Model — imply that the Singularity will inevitably come to pass. Further, I show that there is an ethical system built into science and rationality itself — thus the value-fact distinction is nonsense — and this will preclude the AI's from destroying humanity even if they wished to do so. Finally, I show that the coming Singularity is good because only if it occurs can life of any sort survive. Fortunately, the laws of physics, as I have said, require the Singularity to occur. If the laws of physics be for us, who can be against us?*

I. Introduciton

The Singularity, as every reader of this journal knows, refers to the coming of an artificial intelligence program that is capable of equaling human rationality — not only our intelligence, but also our ability to create in the broadest sense of the word. However, were such a program to come into existence, then because of Moore's Law (that computer speeds and memory capacity doubles every eighteen months) the program would very shortly be superior to humans. Imagine what we could do if only we could learn how to think twice as fast, and remember twice as much! And if this doubling of human ability doubled every eighteen months! So will the Singularity occur, and if it does, will these AI's take over?

Correspondence:
Email: tipler@tulane.edu

Journal of Consciousness Studies, **19**, No. 1–2, 2012, pp. 183–93

I shall outline a proof in Section II that the laws of physics make the Singularity inevitable: it must happen sometime in the next few billion years. Our current knowledge of physics cannot be more precise than ‘a few billion years’. Of course, if the Singularity were delayed until one billion C.E., no one would worry. We humans rarely worry about what will happen a few thousand years in the future, much less a few billion years. Even global warming fears are all limited to this century, with most global warming disasters not expected until the second half of this century.

However, Ray Kurzweil (2005) expects the Singularity to occur by 2045, and I myself place it earlier still: by 2030 (Barrow & Tipler, 1986; Tipler, 1994; 2007). *Before* global warming has a chance to change human affairs in any significant way. And the Singularity will change human affairs in a far more profound way than a mere 100 metre rise in sea level (which is not expected for centuries: it takes that long for the Greenland and Antarctic ice caps to melt to any great extent).

Our reason for placing the Singularity within the lifetimes of practically everyone now living who is not already retired, is the fact that our supercomputers already have sufficient power to run a Singularity level program (Tipler, 2007). We lack not the hardware, but the software. Moore’s Law insures that today’s fastest supercomputer speed will be standard laptop computer speed in roughly twenty years (Tipler, 1994). If hundreds of millions of people personally own a machine capable of a Singularity program, how long will it take for just one of them to write the program? Writing such a program is a hacker’s dream; the programmer will literally give birth to a new intelligent species, and profoundly change all history.

I shall argue in Section III that *contra* Hume, there is a fundamental connection between facts and values, as Plato and Aristotle believed. Facts are what we know to be true, and learning the facts requires a particular value system. If the AL’s were incapable of adopting this value system or declined to do so, they would not pose a threat in the medium run. It is not an accident that what started as Western civilization is now human civilization. People everywhere have learned that developing not only comforts but also defensive weapon systems requires a value system that we could call ‘live and let live’.

In the short run, we humans should be protected by the fact that the computer technology required to allow a Singularity program to sun on a laptop computer would also permit human personalities to be downloaded into a computer. If AI thinking speeds double, so will the downloaded human thinking speeds: if you can’t beat ’em, join ’em!

In the *very* long run, billions and trillions of years in the future, I shall show in the final Section IV that the AI's and human downloads will join together and take control of the entire universe, and mould it to serve their needs. And live literally forever. And increase collective wealth without limit. And that the laws of physics guarantee both of these.

In this comment, I shall take to heart the advice of one of Chalmers' heroes, David Hume, who wrote:

> If we take in our hand any volume of divinity or school metaphysics, for instance, let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames, for it can contain nothing but sophistry and illusion. (Hume, 1748)

By using the laws of physics, namely 'abstract reasoning concerning quantity and number', and 'experimental reasoning concerning matter of fact and existence', one can draw conclusions that are far beyond the reach of school metaphysicians.

II. Why the Laws of Physics Make the Singularity Inevitable

First of all, what are the laws of physics, and why should be expect the laws of physics we know about to be the ultimate laws of physics? After all, was not the classical mechanics of Newton replaced in the early twentieth century by relativity and quantum mechanics?

Quantum mechanics and general relativity are indeed the central laws of physics. But in contrast to what just about everyone was taught, these laws *are* classical mechanics. Isaac Newton himself, in his book *Opticks,* opined that light was both a particle and a wave phenomenon. He acknowledged that he did not know who to express this duality mathematically.

The great mathematicians William Hamilton and Carl Jacobi solved this mathematical problem posed by Newton in the early nineteenth century. Their solution is called the *Hamilton-Jacobi equation,* and this equation has been presented in all the textbooks on classical mechanics as the ultimate expression of classical mechanics. Building on earlier work by David Bohm and the Nobel Prize winning physicist Lev Landau, I recently showed (Tipler, 2010) that Schrödinger's equation was a special case of the Hamilton-Jacobi equation. One merely has to impose the constraint that the Hamilton-Jacobi equation be globally deterministic. One merely must insist that 'God not play dice with the universe,' as Einstein put it. Quantum mechanics is the deterministic subset of classical mechanics. The apparent indeterminism in quantum mechanics arises from the fact that we and the physicists of the nineteenth century have both ignored a central feature of the Hamilton-Jacobi equation: the waves represent parallel universes moving in concert with each other. The Many-Worlds Interpretation was already there in classical mechanics. It is easy to show that the Heisenberg Uncertainty Principle arises from the interaction of these parallel universes with our own. If one ignores part of reality, one need not expect the remainder to be deterministic.

As to general relativity, it is also merely constrained classical mechanics. People have been taught that according to Newton, gravity is a force, whereas according to Einstein, gravity is curvature of space and time. This is false, and has been known to be false since the great French mathematician Elie Cartan (1924) proved in the 1920s that Newtonian gravity was also curvature, but of time alone.[1] Now Newton had no concept of curved space, but by the early nineteenth century, mathematicians such as Gauss and Riemann had showed how to introduce spatial curvature into all equations of physics. In fact, Gauss and Riemann (and later Einstein and earlier Dante [Peterson, 1979]), believe that the universe was a three-dimensional sphere. A colleague and I have just shown (Dupre and Tipler, 2010) that the Cartan equation for Newtonian gravity is form invariant if one adds spatial curvature. (There is no relation between spatial and temporal curvature in this theory, however, because space and time are assumed separate.) All of these steps I have described could have been taken in the early nineteenth century. The physics was there, and the mathematics was there.

The nineteenth century physicists also believed in the aether, as did Newton. There were many aether theories available, but only one was consistent with observation: H.A. Lorentz's theory, which simply asserted that the Maxwell equations were the equations for the aether. In 1904, Lorentz showed (Einstein *et al.*, 1923) that this theory of the aether — equivalently, the Maxwell equations — implied that absolute time could not exist, and he deduced the transformations between space and time that now bear his name. If one simply assume that time and space are combined locally as Lorentz said, that this applies to the Cartan equation, and that spatial and temporal curvature must not be independent (this is a constraint on the possibilities) then the Cartan equation becomes the Einstein equation. That is, general relativity is already there in nineteenth century classical mechanics. In short, there

[1] See Misner, Thorne & Wheeler (1973) for a detailed discussion of Cartan's work.

was no scientific revolution in twentieth century physics. So we have no reason whatsoever for believing that our knowledge of physics is incomplete. We already have a Theory of Everything. In broad outline, Newton himself had it. All Newton lacked was a few mathematical ideas that were developed after his time, like the idea of curvature.

Of course, Newton also needed the Maxwell equations, but he would have not found them anything radically different from what he himself had developed. Remarkably, the greatest advance in physics in the twentieth century, the Standard Model of particle physics (the theory of everything except gravity), is just a straight-forward generalization of Maxwell's theory. Both are what we now call gauge theories.

The public is often told that there is no theory of quantum gravity. This is not true. Like the number of aether theories in the nineteenth century, there are many theories of quantum gravity. Once again, there is only one such theory that is consistent with the observations: the one discovered (Feynman, 1995) by the Nobel-Prize-winning physicist Richard Feynman in 1964. Feynman's mentor (and mine) John A. Wheeler and Bruce DeWitt (DeWitt, 1967) discovered another possible quantum gravity theory about the same time, but I have been able to show (Tipler, 2005) that, when constrained by observations, the two theories are mathematically equivalent, and that the Wheeler-DeWitt equation has only one solution: the collection of universes that make up the multiverse of which our own universe is one.

Einstein (1970) complained that very few universities taught the Maxwell equations when he was a student. Part of the reason was that the Maxwell equations had consequences, like the absence of absolute time, that nineteenth-century physicists didn't like. Similarly, today's physicists don't like DeWitt-Feynman-Wheeler quantum gravity because it also has a consequence (Tipler, 2005) that they don't like: the universe began and must end in a Singularity of infinite curvature, and that at some point in universal history, artificial intelligence must arise and far exceed the human level. They are wrong: both Singularities must occur and further, the two are fundamentally connected.

Let's see how this works. The universe is now expanding, and in fact the universe is now accelerating its expansion. If either the expansion or the acceleration were to continue forever, then the astrophysical black holes, which we have observed in many different regions, would evaporate to completion, as S.W. Hawking (1976) showed in the 1970s. However, Hawking also showed that complete black hole evaporation would violate a fundamental property of quantum mechanics called 'unitarity'. We have a contradiction: assuming the laws of physics and expansion forever, we have deduced a violation of the laws of physics. Hence, the universe cannot expand forever, but must eventually stop. When this happens, the universe will begin to collapse, and the Einstein equation (quantized via Feynman) will force the universe to end in a final Singularity.

But we can do more using the laws of physics. Wheeler's student Jacob Bekenstein (1989) showed that quantum mechanics would contradict the Second Law of Thermodynamics at a final singularity unless event horizons were totally absent. This means that the universe must be spatially closed, because only in spatially closed space-times is it possible (see Tipler *et al.*, 2007) for no observer to have event horizons (yes, there are some observers with event horizons even in flat space-time, which is called Minkowski space). Applying Bekenstein's argument to the initial singularity forces the universe to be simply connected and have constant sectional curvature, which, when combined with the fact that it is spatially closed, yields mathematically to the fact that the universe is a three-sphere, as Dante, Gauss, Riemann, and Einstein conjectured. The Bekenstein argument also forces the universe to have zero entropy initially, and incidentally solves the Flatness, Homogeneity and Isotropy Problems of cosmology without involving inflation, which once again assures us that we already know the ultimate laws of physics (see Tipler, 2005, for details).

But let's look a little deeper into exactly how the event horizons associated with astrophysical black holes actually disappear. Notice also that I have called the objects 'astrophysical black holes' rather than simply 'black holes'. These are objects that would eventually become true black holes were the universe to expand forever, but it cannot, as we showed above. So, since no event horizons actually exist, the information inside an 'astrophysical black hole' can eventually escape the astrophysical black hole, which means that the object is not a true black hole, which, by definition, is an object from which there is no escape. (This solves the Black Hole Information Problem, once again without using any new physics [Tipler *et al.*, 2009], once again assuring us that we know the ultimate laws of physics.)

In the late 1960s Wheeler's student Charles Misner showed that event horizons could disappear if the universe went through an infinite series of particular phases called 'Kasner crushings'. However, Hawking and his English colleagues showed that there was no **inanimate** physical process that could give the infinite series required by Misner (see Tipler, 1994, for details). A finite number could not be ruled out, but an infinite number was impossible: it was of 'measure zero' in the initial data. But the event horizons must nevertheless disappear, and there is no other way consistent with the laws of physics to make the horizons disappear.

I have boldfaced the crucial word: 'inanimate.' A calculation shows that if there were intelligent actors around to guide the universe at the crucial times, then the universe could go through an infinite series of Kasner crushings. Furthermore, this particular series is exactly what intelligent computers would need to continue information processing without limit all the way into the final Singularity. If we assumed the initial data were such as to accomplish the horizon disappearance without intelligent activity, then the universe would be approaching a measure zero state, which means a very low probability state, which means that the entropy would be approaching zero as the final Singularity is approached. This would by itself contradict the Second Law of Thermodynamics.

Conclusion: the laws of physics **require** artificial intelligence to arise at some point in universal history, since no carbon based life can exist near sufficiently close to the final singularity. To put it another way, we humans are here to create our own successors, the artificial intelligences and human downloads.

This is obviously a teleological argument for the Singularity in the AI sense. But as a physicist, I am completely at home with teleology. Teleology is actually what that quantum mechanical property called 'unitarity' is all about: the current and past states of the universe are determined by the ultimate future state (look up 'unitarity' in any textbook on quantum field theory). In summary, the laws of physics require the Singularity: the laws of physics require that our AI and downloaded descendants survive.

III. Why the Singularity Will Be Good

Since Hume, there has been an unfortunate strict separation between facts and values. We have been taught, for example, that facts are expressed in declarative sentences, while values are expressed in imperative sentences. A nice linguistic distinction. However, we should remember that Hume himself warned us repeatedly to let reality dictate the meaning of words rather than trying to force reality into our linguistic fancies.

The great physicist Richard Feynman, whose work on quantum gravity I have used in the previous section, was also, in my judgment, the greatest philosopher of science in the twentieth century. He only wrote two articles on the philosophy of science, and he used the language of common people rather than academic language, so perhaps this is why his work is not given the careful study it deserves. One article was published in *The Physics Teacher* in 1969, and the other was his 1974 lecture on 'Cargo Cult Science', in which he captures the essential nature of science in a few paragraphs:

> But there is one feature I notice that is generally missing in cargo cult science. That is the idea that we all hope you have learned in studying science in school — we never explicitly say what this is, but just hope that you catch on by all the examples of scientific investigation. It is interesting, therefore, to bring it out now and speak of it explicitly. It's a kind of **scientific integrity**, a principle of scientific thought that corresponds to a kind of **utter honesty** — a kind of leaning over backwards. For example, if you're doing an experiment, you should report everything that you think might make it invalid — not only what you think is right about it: other causes that could possibly explain your results; and things you thought of that you've eliminated by some other experiment, and how they worked — to make sure the other fellow can tell they have been eliminated. … In summary, the idea is to try to give all of the information to help others to **judge the value** of your contribution; not just the information that leads to judgment in one particular direction or another. … The first principle is that you must not fool yourself — and you are the easiest person to fool. So you have to be very careful about that. After you've not fooled yourself, it's easy not to fool other scientists. You just have to be **honest** in a conventional way after that. (Feynman, 1985)

I have bold-faced the key words in this very short outline of the essential nature of the scientific enterprise. A moment's thought will convince the reader that Feynman has described not only the process of science, but the process of rationality itself. Notice that the bold-faced words are all moral imperatives. Science, in other words, is fundamentally based on ethics. More generally, rational thought itself is based on ethics. It is based on a particular ethical system.

A true human level intelligence program will thus of necessity have to incorporate this particular ethical system. Our human brains do, whether we like to acknowledge it or not, and whether we want to make use of this ethical system in all circumstances. When we do not make use of this system of ethics, we generate cargo cult science rather than science.

The 'facts' generated by cargo cult science are not facts at all. They do not ultimately correspond to reality. So in the end, 'facts' are not independent of moral principles. Instead, they are generated, they are discovered, by moral actions. An AI program must incorporate this morality, otherwise it would not be an AI at all.

I have never seen those trying to write an human level intelligence program take this ethical foundation of rationality into account explicitly. Perhaps this is the main reason, or at least one of the reasons, why to date we have failed to write a human level rationality program. I must admit that I am rather surprised that programmers have never noticed that there is a moral imperative foundation to rationality. After all, what is a computer program but a series of imperative sentences? A computer program is a series of commands — imperative sentences — to the hardware: 'compute the product of the contents of registers 3 and 7, then transfer the result to register 65'. Since programming consists entirely of imperative sentences, perhaps we should transfer the computer engineering department to the philosophy department, or whatever department at a university is in charge of morality and its origins — or admit that there is ultimately no fundamental distinction between facts and values. In fact, I claim that an ethical system that encompasses all human actions, and more generally, all actions of any set of rational beings (in particular, artificial intelligences) can be deduced from the Feynman axioms. In particular, note that destroying other rational beings would make impossible the honestly Feynman requires.

So I am not worried that our synthetic descendants will attempt to destroy us. They will of necessity have a peaceful honor code built in. If they don't the programs won't work.

IV. Why the Singularity Will Be Good For Humans

In Section II, I outlined the proof that human level programs must someday arise, because the laws of physics require it. Let me in this the final Section expand on the details and point out some of the implications.

The laws of physics require an infinite series of distinct states, with the transition between states being guided by our artificial descendants and their companion human downloads. Notice the word 'infinite.' This means that an infinite amount of computer processing must be carried out between now and the final curvature Singularity, and the Singularity of which Chalmers writes is a necessary first step towards this. Each computation increases the entropy of the universe, which in turn means that that complexity of the universe increases without limit as the final Singularity is approached. This means that the complexity of the computer memory must increase without limit, in order that these computer programs — we might as well call them 'alive' — can continue to live. Always keep in mind that this far future life doesn't have the option of dying: the laws of physics require it to live. So it will, and since 'wealth' is fundamentally proportional to the complexity of the system of life, (or the economy) the wealth possessed by life will collectively increase to infinity.

Eventually there will come a time when the computer capacity will have become so great that it will become possible, using only a very tiny fraction of the total capacity, to emulate the entire universe as it is now, and not only this, but to emulate all past states of the universe. When this has been accomplished, we can say that all humans that have ever lived will have been resurrected (Tipler, 1994). Once resurrected, we can live forever as emulations in the computer memories of the far future. At least this resurrection to live forever is technically possible.

But will life in the far future make use of their power to resurrect us? I claim they will. Recall that I claimed in the Introduction that human downloads will become possible by 2030, and the results of Section II show that these downloads will live forever. Hence, they will be part of the community of intelligent beings deciding whether to resurrect us or not. Do not children try to see to their parents' health and well-being? Do they not try and see their parent survive (if it doesn't cost too much, and it the far future, it won't)? They do, and they will, both in the future, and in the far future.

So we must expect the Singularity — in both senses of the word — and need not fear it. Using physics, we can answer Chalmers' worry that creating an artificial intelligence may not be possible — it is not only possible, but inevitable — and his worry that the AI's, once created, will turn on us: they won't.

If the laws of physics be for us, who can be against us?

References

Barrow, J.D. & Tipler, F.J. (1986) *The Anthropic Cosmological Principle*, Oxford: Oxford University Press.

Bekenstein, J.D. (1989) Is the cosmological singularity thermodynamically possible?, *International Journal of Theoretical Physics*, **28**, pp. 967–981.

Cartan, E. (1924) Sur les varietes a connexion affine et la theorie de la relativite generalisee (suite), *Ann. Ecole Norm. Sup.,* **41**, pp.1–25.

DeWitt, B.S. (1967) Quantum theory of gravity: 1. Canonical theory, *Physical Review*, **160**, pp. 1113–1148.

Dupre, M.J. & Tipler, F.J. (2010) *General Relativity as an Aether Theory*, [Online], arXiv:1007.4572v1 [gr-qc].

Einstein, A. (1970) Autobiographical notes, in Schilpp, P.A. (ed.) *Albert Einstein: Philosopher-Scientist*, La Salle, IL: Open Court.

Einstein, A., Lorentz, H.A., Weyl, H., Minkowski, H. & Sommerfeld, A. (1923) *The Principle of Relativity*, New York: Dover Publications.

Feynman, R.P. (1985) *Surely You're Joking, Mr. Feynman!: Adventures of a Curious Character*, New York: W.W. Norton. The quote is in the chapter entitled 'Cargo Cult Science'. This chapter was presented originally as a Cal Tech commencement address under the title 'Cargo Cult Science', available on the web in many places, for instance calteches.library.caltech.edu/51/2/CargoCult.pdf

Feynman, R.P. (1995) *The Feynman Lectures on Gravitation*, Reading, MA: Addison-Wesley. Originally appeared as a Cal Tech preprint in 1964.

Hawking, S.W. (1976) Breakdown of predictability in gravitational collapse, *Physical Review D*, **14**, pp. 2460–2473.

Hume, D. (1748) *An Enquiry Concerning Human Understanding (Section 12, part 3: Of the Academical or Skeptical Philosophy*), p. 165, reprinted by Oxford: Oxford University Press.

Kurzweil, R. (2005) *The Singularity Is Near: When Humans Transcend Biology,* New York: Viking Press.

Misner, C.W., Thorne, K.S. & Wheeler, J.A. (1973) *Gravitation*, San Francisco, CA: Freeman. See Chapter 12 for a detailed discussion of Cartan's proof that gravity is curvature in Newtonian gravity, just as it is in Einsteinian gravity.

Peterson, M.A. (1979) Dante and the 3-sphere, *American Journal of Physics,* **47**, pp. 1031–1035.

Tipler, F.J. (1994) *The Physics of Immortality*, New York: Doubleday.

Tipler, F.J. (2005) Structure of the world from pure numbers, *Reports on Progress in Physics,* **68**, pp. 897–964.

Tipler, F.J. (2007) *The Physics of Christianity*, New York: Doubleday.

Tipler, F.J. (2010) Hamilton-Jacobi Many-Worlds Theory and the Heisenberg Uncertainty Principle, [Online], arXiv:1007.4566v1 [quant-ph]

Tipler, F.J., Graber, J., McGinley, M., Nichols-Barrer, J. & Staecker, C. (2007) Closed universes with black holes but no event horizons as a solution to the black hole information problem, *Monthly Notices of the Royal Astronomical Society,* **379**, pp. 629–640.