

In: Strategy for the Future (Bushko R, ed.), 2008, in press.

The singularity and the Methuselarity: similarities and differences

Aubrey D.N.J. de Grey, Ph.D.

Methuselah Foundation

PO Box 1143, Lorton, VA 22079, USA

Email: aubrey@sens.org

Abstract

Aging, being a composite of innumerable types of molecular and cellular decay, will be defeated incrementally. I have for some time predicted that this succession of advances will feature a threshold, which I here christen the “Methuselarity,” following which there will actually be a progressive decline in the rate of improvement in our anti-aging technology that is required to prevent a rise in our risk of death from age-related causes as we become chronologically older. Various commentators have observed the similarity of this prediction to that made by Good, Vinge, Kurzweil and others concerning technology in general (and, in particular, computer technology), which they have termed the “singularity.” In this essay I compare and contrast these two concepts.

The singularity: a uniquely unique event in humanity’s future

“Unique” is, of course, an over-used word to describe momentous events – arguably, even more over-used than “historic.” How, then, can I dare to describe something as uniquely unique?

Well, I will begin by pulling back a fraction from that description. There are actually, in my view, two possible events in humanity’s future that merit this description. But I do not feel very bad about this qualification, because I believe that those two events are, in all probability, mutually exclusive. The singularity is one; the demise of humanity is the other. Hence my choice of the indefinite article: the singularity is not “the” uniquely unique event in humanity’s future, because it may not occur, but if it does occur, nothing comparable will either precede or follow it.

The singularity has been defined in many related but subtly distinct ways over the years, so let me begin my discussion of it by making clear what I mean by the term. I adhere to the following definition: “an asymptotically rapid increase in the sophistication of technology on whose behaviour humans depend.” I do not use the word to mean, for example, “the technological creation of smarter-than-human intelligence” (which is the definition currently given by SIAI, the Singularity Institute for Artificial Intelligence¹) – despite my agreement with the view that the technology most likely to bring about the singularity (and, indeed, the one that was originally used to define it) is precisely the one that SIAI study, namely recursively self-improving artificial intelligence (of which more below). I am sticking to the more abstract definition partly because it seems to me to encapsulate the main point of why the singularity is indeed uniquely unique, and partly because it will help me to highlight what distinguishes the singularity from the Methuselarity.

One aspect of my definition that may raise eyebrows is its use of the word “asymptotically” rather than “exponentially.” I feel sure that von Neumann² would agree with me on this: the mere perpetuation of Moore’s Law³ will not bring about the singularity. A gravitational singularity, which is of course the etymological source of the term, is the centre (not, I stress, the event horizon) of a black hole: the point at which the force of gravity is infinite – or, to be more precise, the point arbitrarily near to which

gravity is arbitrarily strong. The distance between the singularity and any point of interest (inside or outside the event horizon) at which gravity is finite is, of course, finite. This is an asymptotic relation between distance and strength: if point X is distance Y from the singularity, it is not possible to travel from X, along the line between X and the singularity, by a distance greater than Y, and experience continuously increasing gravity. Exponential (though not inverse exponential! – see below) relations are not like this: they have no asymptote. If the force of gravity exerted by a particular body were exponential (though still increasing with decreasing distance from the body), the relation between distance from that body and gravity exerted by it would be defined in terms of distance from the point furthest away from it (“on the other side of the Universe”). Call the gravity exerted at that point X and suppose that the gravity exerted at half that distance from the body is 4X (which is the same as for gravity in real life). Then the gravity exerted by the body at a point arbitrarily close to it is not arbitrarily large – it is just 16X, since that point is exactly twice as far away from the point of minimum gravity as the 4X point is.

Having belaboured this point, I now hope to justify doing so. Will the technological singularity, defined as I define it above, happen at all? Not if we merely proceed according to Moore’s law, because that does not predict infinite rates of progress at any point in the future. But wait – who’s to say that progress will remain “only” exponential? Might not progress exceed this rate, following an inverse polynomial curve (like gravity) or even an inverse exponential curve? I, for one, don’t see why it shouldn’t. If we consider specifically the means whereby the Singularity is most widely expected to occur, namely the development of computers with the capacity for recursive improvement of their own workings,⁴ I can see no argument why the rate at which such a computer would improve itself should not follow an inverse exponential curve, i.e. one in which the time taken to achieve a given degree of improvement takes time X, the time taken to repeat that degree of improvement is X/2, then X/4 and so on.

Why does this matter? It might matter quite a lot, given that (in most people’s view, anyway) the purpose of creating computers that are smarter than us is to benefit us rather than to supersede us. Human intelligence, I believe, will not exhibit a super-exponential rate of growth, because our cognitive hardware is incompatible with that. Now, I grant that I have only rather wishy-washy intuitive reasons for this view – but what I think can be quite safely said is that our ability to “keep up” with the rate of progress of recursively self-improving computers will be in inverse relation to that rate, and thus that super-exponentially self-improving computers will be more likely to escape our control than “merely” exponentially self-improving ones will. Computers have hardware constraints too, of course, so the formal asymptotic limit of truly infinite rates of improvement (and, thus, truly infinite intelligence of such machines) will not be reached – but that is scant solace for those of us who have been superseded (which could, of course, mean “eliminated”) some time previously. There is, of course, the distinct possibility that even exponentially self-improving systems would similarly supersede us, but the work of SIAI and others to prevent this must be taken into account in quantifying that risk.

Let us now consider the aftermath of a “successful” singularity, i.e. one in which recursively self-improving systems exist and have duly improved themselves out of sight, but have been built in such a way that they permanently remain “friendly” to us. It is legitimate to wonder what would happen next, albeit that to do so is in defiance of Vinge.⁵ While very little can confidently be said, I feel able to make one prediction: that our electronic guardians and minions will not be making their superintelligence terribly conspicuous to us. If we can define “friendly AI” as AI that permits us as a species to follow our preferred, presumably familiarly dawdling, trajectory of progress, and yet also to maintain our self-image, it will probably do the overwhelming majority of its work in the background, mysteriously keeping things the way we want them without worrying us about how it’s doing it. We may dimly notice the statistically implausible occurrence of hurricanes only in entirely unpopulated regions, of sufficiently deep snow in just the right places to save the lives of reckless mountaineers, and so on – but we will not dwell on it, and quite soon we will take it for granted.

A reasonable question to ask is, well, since even a super-exponentially self-improving AI will always have finite intelligence, might it not at some point create an even more rapidly self-improving system that could supersede it? Indeed it might (I think) – but, from our point of view, so what? If we have succeeded in creating a permanently friendly AI, we can be sure that any “next-generation” AI that it created would also be friendly, and thus (by the previous paragraph’s logic) largely invisible. Thus, from our perspective, there will only be one singularity.

In closing this section I return to my claim that the singularity and the demise of humanity are, in all probability, mutually exclusive. Clearly if our demise precedes the singularity then the singularity cannot occur. Can our demise occur if preceded by the singularity? Almost certainly not, I would say: the interval available for our demise between the development of recursively self-improving AI and the attainment by that AI of extremely thorough ability to protect us (even from, for example, nearby supernovae) will be short. (I exclude here the possibility that the singularity will occur via the creation of AI that is not friendly to us, only because I think humanity’s life expectancy in that scenario is so very short that this is equivalent from our point of view to the singularity not occurring at all.) The “area under the curve” of humanity’s probability of elimination at any time after the singularity is thus very small. I am, of course, discounting here the possibility that even arbitrarily intelligent and powerful systems cannot protect us from truly cosmic events such as the heat death of the Universe, but I agree with Deutsch⁶ that this is unlikely given the time available.

The Methuselararity: the biogerontological counterpart of the singularity

In a recent interview, Watson was asked what would be the next event in the history of biology that would compare in significance to his and Crick’s discovery of the structure of DNA, and he replied that there would never be one.⁷ I think he was correct. However, I agree with him only if I am rather careful in defining “biology” as the *discovery* of features of the living world, and excluding biotechnology, which for present purposes I define as the *exploitation* of such discoveries. In biotechnology I believe that there will certainly be a counterpart, something that will outstrip in significance every other advance either predating or following it: the Methuselararity.

For almost a decade following my graduation in 1985, I conducted research in artificial intelligence. I switched fields to biogerontology shortly after becoming aware that the defeat of aging was vastly less on biologists’ agenda than I had hitherto presumed. I was not, at that time, aware of the concept of recursively self-improving AI and the singularity, though perhaps I should have been. But even if I had been, I think I would still have made the career change that I did. Why?

Humans are very, very good at adjusting their aspirations to match their expectations. When things get better, people are happy – but if they stay better and show every sign of continuing that way, people become blasé. Conversely, when things get worse people are unhappy, but if they stay worse and show every sign of continuing that way, people become philosophical. This is why, by all measures that have to my knowledge been employed, people in the developed world are on average neither much happier nor much less happy now than they were when things were objectively far worse. This is a good thing in many ways, but in at least one way it is a problem: it dampens our ardour to improve our lives more rapidly. In particular, it depletes the ranks of “unreasonable men” to whom Shaw so astutely credited all progress.⁸ There are far too few unreasonable men and women in biology, and especially in biogerontology. I am proud to call myself an exception: someone who is comfortable devoting his life to the most important problems of all, even if they appear thoroughly intractable.⁹ In my youth, I felt I could make the most difference to the world by helping to develop intelligent computers; but when I discovered the truth about biologists’ attitude to aging I knew that I could make even more difference in that field.

Why is aging so important? Aging kills people, yes, but so do quite a few other things – and moreover, life is about quality as well as quantity, and intelligent machines might very greatly improve the quality of life of an awful lot of people, not least by virtue of providing essentially unbounded prosperity for all.

Even if we take into account the fact that aspirations track expectations, such that what really matters is to maintain a good *rate of improvement* of (objective) quality of life, it is hard to deny that the development of super-intelligent machines will be of astronomical benefit to our lives. But let's be clear: quantity of life matters too. There is a well-established metric that folds together the quality and quantity benefits of a given technological or other opportunity: it is the "quality-adjusted life year" or QALY.¹⁰

Historically, mainstream biogerontologists have been publicly cautious regarding predictions of the biomedical consequences of their work, though this is gradually changing. But even privately, few biogerontologists have viewed aging as amenable to dramatic change: they have been aware that it is a hugely multi-faceted phenomenon, which will yield only incrementally to medical progress if it yields at all. This places them in a difficult position when arguing for the importance of their work relative to other supplicants for biomedical research resources. Yes, there is always a benefit to a QALY, and yes, progress against aging will deliver QALYs – but the force of this argument is diminished by two key factors, namely the probability of success (which biogerontologists cannot provide a conclusive case for being high) and the entrenched ageism in society, which views it as "fair" to deprioritise health care for the elderly. This quandary is well illustrated by the current "Longevity Dividend" initiative, which seeks to focus policy-makers' minds on the ever-dependable lure of lucre associated with keeping people youthful, rather than on the moral imperative.¹¹

But this is in the process of changing – indeed, of being turned on its head. This is for one reason and one only: it is becoming appreciated that aging may be amenable to comprehensive postponement by regenerative medicine.^{12,13} And the reason that makes all the difference is because it creates the possibility – indeed, the virtual certainty – of the Methuselarity.

Having tantalised you for so long, I cannot further delay revealing what the Methuselarity actually is. It is the point in our progress against aging at which our rational expectation of the age to which we can expect to live without age-related physiological and cognitive decline goes from the low three digits to infinite. And my use here of the word "point" is almost accurate: this transition will, in my view, take no longer than a few years. Hence the – superficial – similarity to the singularity.

I have set out elsewhere, first qualitatively¹⁴ and then quantitatively,¹⁵ the details of my reasons for believing that the application of regenerative medicine to aging will deliver this cusp; thus, here I will only summarise. Regenerative medicine, by definition, is the partial or complete restoration of a damaged biological structure to its pre-damaged state. Since aging is the accumulation of damage, it is in theory a legitimate target of regenerative medicine, and success in such a venture would constitute bona fide rejuvenation, the restoration of a lower biological age. (The bulk of my work over the past decade can be summarised as the elaboration of that "theory" into an increasingly detailed and promising project plan for actual implementation¹⁶ – but I digress.) This rejuvenation would not be total: some aspects of the damage that constitutes aging would be resistant to these therapies. But not intrinsically resistant: all such damage could in principle be reversed or obviated by sufficiently sophisticated repair-and-maintenance (i.e., regenerative) interventions. Thus arises the concept of a rate of improvement of the comprehensiveness of these rejuvenation therapies that is sufficient to outrun the problem: to deplete the levels of all types of damage more rapidly than they are accumulating, even though intrinsically the damage still present will be progressively more recalcitrant. I have named this required rate of improvement "longevity escape velocity" or LEV.^{14,15}

It is important to understand that LEV is not an unchanging quantity, as it might be if it were a feature of our biology. Rather, it will vary with time – and exactly how it will probably vary is a topic I address in the next section. LEV will, however, remain non-zero for as long as there remain any types of damage that we cannot remove or obviate. Thus, the formal possibility exists that we will at some point achieve LEV but that at some subsequent date our rate of progress against aging will slip back below LEV. However, I have claimed that this will almost certainly not happen: that, once surpassed, LEV will be

maintained indefinitely. This claim is essentially equivalent to the claim that the Methuselararity will occur at all: the Methuselararity is, simply, the one and only point in the future at which LEV is achieved.

The singularity and the Methuselararity: some key differences

Having described the singularity and the Methuselararity individually, I now examine how they differ. I hope to communicate that the superficial similarities that they exhibit evaporate rather thoroughly when one delves more deeply.

Perhaps the most important contrast between the singularity and the Methuselararity is the relevance of accelerating change. In the first section of this essay I dealt at some length with the range of trajectories that I think are plausible for the rate of improvement of self-improving artificial intelligence systems – but it will have been apparent that all the trajectories I discussed were accelerating. It might intuitively be presumed that, since aging is a composite of innumerable types of damage that accumulate at different rates and that possess different degrees of difficulty to remove, our efforts to maintain youth in the face of increasing chronological age will require an accelerating rate of progress in our biomedical prowess. But this is not correct.

The central reason why progress need not accelerate is that there is a spectrum not only in the recalcitrance of the various types of damage that constitute aging but also in their rates of accumulation. As biomedical gerontologists, we will always focus on the highest-priority types of damage, the types that are most in danger of killing people. Thus, the most rapidly-accumulating types of damage will preferentially be those against which we most rapidly develop repair-and-maintenance interventions. There will, to be sure, be “spikes” in this distribution – types of damage that accumulate *relatively* rapidly and are also *relatively* hard to combat. But we are discussing probabilities here, and if we aggregate the probability distributions of the timeframes on which the various types of damage, with their particular rates of accumulation and degrees of difficulty to combat, are in fact brought under control, the conclusion is clear: we are almost certain to see a progressive and unbroken *decline* in the rate at which we need to develop new anti-aging therapies once LEV is first achieved. (I do not mean to say that this progression will be absolutely monotonic – but the “wobble” in how rapidly progress needs to occur will be small compared to the margin of error available, i.e. the margin by which the average rate of progress exceeds LEV.) This conclusion is, of course, subject to assumptions concerning the distribution of these types of damage on those two dimensions – but, in the absence of evidence to the contrary, a smooth (log-normal, or similar) distribution must be assumed.

The other fundamental difference between the singularity and the Methuselararity that I wish to highlight is its impact on “the human condition” – on humanity’s experience of the world and its view of itself. I make at this point perhaps my most controversial claim in this essay: that in this regard, the Methuselararity will probably be far *more* momentous than the singularity.

How can this be? Surely I have just shown that the Methuselararity will be the consequence of only quite modest (and, thereafter, actually decreasing) rates of progress in postponing aging, whereas the singularity will result from what for practical purposes can be regarded as infinite rates of progress in the prowess of computers? Indeed I have. But when we focus on humanity’s experience of the world and its view of itself, what matters is not how rapidly things are changing but how rapidly those changes affect us. In the case of the singularity, I have noted earlier in this essay that if we survive it at all (by virtue of having succeeded in making these ultra-powerful computers permanently friendly to us) then we will move from a shortly-pre-singularity situation in which computers already make our lives rather easy to a situation in which they fade into the background and stay there. I contend that, from our point of view, this is really not much of a difference, psychologically or socially: computers are already far easier to use than the first PCs were, and are getting easier all the time, and the main theme of that progression is that we are increasingly able to treat them as if they were not computers at all. It seems to me that the singularity may well, in this regard, merely be the icing on a cake that will already have been baked.

Compare this to the effect of the Methuselarity on the human condition. In this case we will progressively and smoothly improve our remaining life expectancy as calculated from the rate of accumulation of those types of damage that we cannot yet fix. So far, so boring. But wait – is that the whole story? No, because what will matter is the bottom line, how long people think they’re actually going to live.

These days, people are notoriously bad at predicting how long they’re going to live. There is a strong tendency to expect to live only about as long as one’s parents or grandparents did (just so long as they died of old age, of course).¹⁷ This is clearly absurd, given the rapid rise of life expectancies throughout the developed world in the past half-century and the fact that, unlike the previous half-century, that rise has resulted from falling mortality rates at older ages rather than in infancy or childbirth. It persists, I believe, simply because the rise in life expectancy has been rapid only by historical standards: unless one’s paying attention, it’s not been rapid by the standards of progress in technology, so it easily goes unnoticed.

This will not last, however. As the rate of improvement in life expectancy increases, so the disparity between that headline number and the age which someone of any particular age can expect to reach also increases. But here’s the crux: these two quantities do not increase in proportion. In particular, when the rate of improvement of life expectancy reaches one year per year – which, in case you didn’t know, is only a few times faster than is typical in the developed world today¹⁸ – the age that one can expect to reach undergoes a dramatic shift, because the risk of dying from age-related causes at any given age suddenly plummets to near zero. And that is (another way of defining) the Methuselarity.

To summarise my view, then: the singularity will take us from a point of considerable computing power that is mostly hidden from our concern to one of astronomical computing power that is just slightly more hidden. The Methuselarity, by contrast, will take us from a point of considerable medical prowess that only modestly benefits how long we can reasonably expect to live, to one of just slightly greater medical prowess that allows us confidence that we can live indefinitely. The contrast is rather stark, I think you will agree.

Epilogue: the Methuselarity and the singularity combined

Those who have followed my work since I began publishing in biogerontology may have noticed a subtle change in the way that I typically describe the Methuselarity’s impact on lifespans. Early on, I used to make probabilistic assertions about future life expectancy; now I make assertions about how soon we will see an individual (or a cohort) achieve a given age.

The reasons for this shift are many; some are down to my improved sense of what does and does not scare people. But an important reason is that my original style of prediction incorporated the implicit assumption that the Methuselarity would occur in the context of a continued smooth, and relatively slow, rate of reduction in our risks of death from causes unrelated to our age. I only belatedly realised that this assumption is unjustified – indeed, absurd. And the singularity is what makes it particularly absurd.

Roughly speaking, we prioritise our effort to avoid particular risks of death on the basis of the relative magnitude of those risks. Things that only have a 0.01% risk per year of killing us may not be considered worth working very hard to avoid, because even multiplied up over a long life they have only a 1% chance of being our cause of death. This immediately tells us that such risks will move altogether nearer to the forefront of our concerns as and when the Methuselarity occurs (or is even widely anticipated), because the greater number of years available to get unlucky means that the risk of these things being our cause of death is elevated. It seems clear that we will work to do something about that – to improve the efficiency with which we develop vaccines, to make our cars safer, and so on. But there would appear to be only so much we *can* do in that regard: first of all there are things that we really truly

can't do anything about, such as nearby supernovae, and secondly there are quite a few moderately risky activities that quite a lot of us enjoy.

The singularity changes all that. What the singularity will provide is the very rapid reduction to truly minute levels of the risk of death from any cause. You may have thought that my earlier mention of snow reliably saving careless mountaineers was in jest; indeed it was not. Moreover, the residual risk that our rate of improvement of medical therapies against aging will at some point fall below LEV will also essentially disappear with the singularity. (Clearly the possibility also exists that the singularity will precede, and thus bring about, the Methuselarity – but that does not materially alter these considerations.)

One of my “soundbite” predictions concerning the Methuselarity is that the first thousand-year-old is probably less than 20 years younger than the first 150-year-old. The above considerations lead to a supplementary prediction. I think it is abundantly likely that the first million-year-old is less than a year younger than the first thousand-year-old, and the first billion-year-old probably is too.

The singularity and the Methuselarity are superficially similar, but I hope to have communicated in this essay that they are in fact very different concepts. Where they are most similar, however, is in the magnitude of their impact on humanity. The singularity will be a uniquely dramatic change in the trajectory of humanity's future; the Methuselarity will be a uniquely dramatic change in its perception of its future. Together, they will transform humanity... quite a lot.

References

1. Singularity Institute for Artificial Intelligence. What is the Singularity? <http://singinst.org/overview/whatisthesingularity> (retrieved 25th August 2008).
2. Ulam S. Tribute to John von Neumann. *Bulletin of the American Mathematical Society* 1958; 64(3 part 2): 1-49.
3. Moore GE. Cramming more components onto integrated circuits. *Electronics* 1965; 38(8): no pagination.
4. Kurzweil R. *The Singularity Is Near: When Humans Transcend Biology*. New York: Penguin, 2006 (ISBN: 0143037889).
5. Vinge V. The Coming Technological Singularity. In: *Vision-21: Interdisciplinary Science & Engineering in the Era of CyberSpace*, proceedings of a Symposium held at NASA Lewis Research Center (NASA Conference Publication CP-10129), 1993.
6. Deutsch D. *The fabric of reality*. New York: Penguin, 1998 (ISBN: 014027541X).
7. Weatherall D. Was there life after DNA? *Science* 2000; 289(5479):554-555.
8. Shaw GB. Maxims for Revolutionists. In: *Man and Superman*, 1903.
9. de Grey ADNJ. Long live the unreasonable man. *Rejuvenation Res* 2008; 11(3):541-542.
10. Pliskin JS, Shepard DS, Weinstein MC. Utility Functions for Life Years and Health Status. *Operations Research* 1980; 28:206-224.
11. Olshansky SJ, Perry D, Miller RA, Butler RN. Pursuing the longevity dividend: scientific goals for an aging world. *Ann N Y Acad Sci* 2007; 1114:11-13.
12. de Grey ADNJ, Ames BN, Andersen JK, Bartke A, Campisi J, Heward CB, McCarter RJM, Stock G. Time to talk SENS: critiquing the immutability of human aging. *Annals NY Acad Sci* 2002; 959:452-462.
13. de Grey ADNJ. A strategy for postponing aging indefinitely. *Stud Health Technol Inform* 2005; 118:209-219.

14. de Grey ADNJ. Escape velocity: why the prospect of extreme human life extension matters now. *PLoS Biol* 2004; 2(6):723-726.
15. Phoenix CR, de Grey ADNJ. A model of aging as accumulated damage matches observed mortality patterns and predicts the life-extending effects of prospective interventions. *AGE* 2007; 29(4):133-189.
16. de Grey ADNJ, Rae M. *Ending Aging: The rejuvenation biotechnologies that could reverse human aging in our lifetime*. New York, NY: St. Martin's Press, 2007, 416pp, hardcover (ISBN 0-312-36706-6).
17. Banks J, Emmerson C, Oldfield Z. Not so brief lives: longevity expectations and wellbeing in retirement. In: *Seven Ages of Man and Woman* (Stewart I and Vaitilingam R, eds.), Swindon: Economic and Social Research Council, 2004, pp. 28-31.
18. Oeppen J, Vaupel JW. Broken limits to life expectancy. *Science* 2002;296(5570):1029-1031.