Vajon mi történik, ha egy biológus úgy dönt, hogy saját magát reprodukáló számítógépes programok kis világát teremti meg, hogy vizsgálhassa az evolúciót?

https://en.wikipedia.org/wiki/Tierra_%28computer_simulation%29

https://en.wikipedia.org/wiki/Thomas_S._Ray

http://life.ou.edu/tierra/

Lásd még Lem: Lymphater képlete

William Lethem Organic world

Az evolúciós folyamat modellezéséhez szándékosan el kellett rontani néhány bitet a modellbe betáplált "kód-organizmusokban". A tökéletesen működő számítógép belsejében egyfajta virtuális gépet hozott létre, amelyik kevésbé tökéletesen működik, mondjuk minden ötszázadik bitet hibásan másol át. Ezzel a másolási hibák gyakorisága szabályozott volt. Így alkotta meg Thomas Ray a "mesterséges Föld" absztrakt modelljét. Erre utalva a spanyol TIERRA nevet kapta, ami Földet jelent.

A TIERRA környezetben egy egység 5 bitből áll. Ezzel 32 utasítás volt kódolható. Ray kidolgozta, hogy milyen utasításkészlet alkossa a TIERRA programozási nyelvét. Úgy válogatta össze a 32 utasítást, hogy olyan programot lehessen írni, amik a TIERRA memóriájában létrehozzák önmaguk egy (a már említett random hibákkal létrejövő) másolatát.
Matematikusok később bebizonyították, hogy ez a szélsőségesen lecsupaszított számítógép is valójában éppen olyan univerzális, mint Neumann számítógépe volt. (Neumann megmutatta, hogy az univerzális számítógép képes önmagukat fizikailag reprodukáló robotok készítésére.) Ha megépítenénk valódi gépnek, azon is lehetne futtatni szövegszerkesztő-, játék- vagy böngészőprogramokat is. Ez fontos tény: nem lehet ugyanis azt állítani, hogy a TIERRA csak szimuláció, aminek alig van köze a valódi folyamatokhoz.
A TIERRA környezetben lényegét tekintve számítógépes vírusokhoz hasonló programok futottak. Egymással versenyeztek a processzor erőforrásaiért és a rendelkezésre álló memóriaterületért, amelyek az energia és az életterület absztrakt analógiájának tekinthetők. Ebben a környezetben az a program, amelyik kevesebb helyet foglal el, könnyen talál az utódjának helyet. Az amelyik gyorsabban, kevesebb

processzorművelet felhasználásával szaporodik, szintén előnybe kerülhet.

A TIERRA környezet a benne futó konkrét programoktól függetlenül gondoskodik a másolási hibákról, illetve az idős, valamint a rendszert gyakran lefagyasztó programok elpusztításáról. (Találó neve ezen alrésznek a Reaper [Kaszás] volt.)

"Lehet zöld az ég, és lehet kék a fű..."

Amikor a környezet elkészült, Thomas Ray legjobb tudása szerint írt egy programot, amely képes volt elkészíteni önmaga másolatát. A legrövidebb ilyen program, amit el tudott képzelni, 80 utasításból állt. 1990. január 3-án sikerült először lefuttatnia ezt a programot a frissen elkészült TIERRA virtuális gépen, és mint írja, "ezzel elszabadult a pokol. [...] Kutatási programom egyszerre csak gyökeresen átalakult. Eddig egy rendszer tervezése volt a feladatok, innentől pedig csak a megfigyelés. Újra a dzsungelben voltam, hogy leírjak egy idegen világot, amelynek fizikája és kémiája teljesen másmilyen volt, mint az élet azon formáié, amelyeket ismertem és szerettem. Mégis, az élet olyan formái és folyamatai jelentek meg, amelyek egy gyakorlott természetkutató szemével valahogy nagyon is jól azonosíthatók voltak."
A gyakorlott természetkutató (és a hírszerzési dolgozó is) tudja, hogy amikor így ül megfigyelőhelyén, sokáig nem történik semmi érdekes. De figyelme nem lankadhat, bármikor történhet valami fontos dolog. Ray nem akart hinni a szemének, amikor egy 79 utasításból álló lényt látott, amely valamivel gyorsabban szaporodott, mint az ő 80-as utasítássorozata. Hosszabb programok is megjelentek, de azok idővel kihaltak. Egyszer csak megjelent egy 45 utasításból álló program, ami a vizsgálatok alapján egy valódi parazitának bizonyult: egy másik program kódjának részletét is felhasználta szaporodásához: megkereste az egyik 80-as kódját a memóriában, rákapcsolódott egy alkalmas ponton, és így magát a másolást (mint szaporodást) már ezzel végeztette el.
Számomra itt válik még érdekesebbé a történet. Egyes programok immunitást fejlesztettek ki az őket gyötrő parazita ellen, például a 79-es immunissá vált a 45-ös parazita egy változatára. Ám ekkor megjelent egy 51 soros, ami ezt a védelmet fel tudta törni. Még az altruizmus különböző formái is megjelentek, és hosszabb időn át fenn is maradtak. Előfordult, hogy egy parazita fejlesztett ki védelmet a "gazdaszervezete" számára, és ezzel annak ellenére jól járt, hogy maga egy kicsit hosszabbá vált.
Így történt, hogy Thomas Raynek soha többé nem kellett önreprodukáló programot írni. Gyártott neki rengeteget a TIERRÁ-ban zajló evolúció. Ezzel zárta ki azt a korlátozó tényezőt, ami az evolúció kutatásában állt a tudósok előtt: hogy csak egy van belőle, és azt is csak kis részletekben, régészeti leletekből vizsgálhatjuk. A

TIERRA sokszori újraindítása során mindig más-más ökoszisztéma alakult ki, másfajta élőlényekkel és ezek legkülönbözőbb fajta összefonódásával. Néha még az ivaros szaporodás valamiféle formája is kifejlődött: egy programpár először is megkereste egymást a szaporodáshoz, ezek után mindkettő a TIERRA memóriájának ugyanarra a részére kezdte másolni önmagát, és így az utód által elfoglalt memóriaterület egy részére az egyik, a többi részére pedig a másik program kódrészlete került.
A TIERRÁ-nak sikerült elképesztenie Rayt, mint biológust és mint programozót. Egyszer váratlanul megjelent egy 22 utasításból álló program, amely teljesen egyedül is képes volt szaporodni. Nem volt sem parazita, sem másképpen nem függött más programoktól. Joggal döbbentette meg a program hatékonysága: negyed annyi utasítás is elég volt a kitűzött célhoz. Az evolúció minden megjelenési formájában csodákra képes.

¤¤

Evolúciós algoritmusok
Thomas Ray / Tierra Project

http://index.hu/kultur/cyb/evolucio/

http://life.ou.edu/tierra/

Darwin és az evolúció elmélete talán még soha nem örvendett akkora népszerűségnek, mint napjainkban. Pedig a XX. században több evolucionista divat (például a szociál-darwinizmus) végigsöpört. Igaz, azok általában csak egy-egy - a rendszer egészéből, a kontextusból kiragadott - gondolatot fejlesztettek tovább.
A "vulgár-darwinizmus" még lejjebb ment: a Mester meglátásait, érveit szlogen-szintre süllyesztő kocsmafilozófiáknak se szeri, se száma.

Legizgalmasabb terepnek, a Richard Dawkins munkásságának köszönhetően sokadik reneszánszát élő evolúciós biológia tűnik. A Dawkins műveiből kiinduló poszt-posztmodern (?) evolúció a tudomány szinte valamennyi szférájában meghatározó szerepet játszik. Hans Moravec és Kevin Warwick - egymástól eltérő - robotjövői, a mesterséges intelligencia (AI), vagy a mesterséges élet (A-Life) kutatása elképzelhetetlen nélküle.

Tény az is, hogy a számítógépes világban az evolúciós és a genetikai algoritmusok egyre nagyobb jelentőségre tesznek szert.

Az előbbi kategóriába azokat a (káosz- és komplexitás-elméletekhez szintén köthető) computeres problémamegoldó rendszereket soroljuk, melyek tervezése során az evolúció mechanizmusának számítógépes modellje a központi elem. A (nagyjából azonos) genetikus algoritmus a gépi tanulás azon modellje, mely viselkedésében a természetes evolúció folyamatát utánozza. Egy programon belül, mesterséges individuumokból álló "populációt" hozunk létre, és ezek - kompetitív, szimulált környezetben - az evolúció mechanizmusának "vetik alá magukat".

S hogy miért pont ezek az algoritmusok lehetnek sikeresek? Induljunk ki a biológiai modellből: a természetes kiválasztódáson alapuló evolúció komplex és intelligens létformákat generál. Sőt: egyre komplexebbeket és egyre intelligensebbeket. Továbbá, ellentétben a teremtés-elméletekkel, az evolúció az egyetlen olyan folyamat, mely bizonyíthatóan intelligenciát "hoz világra".

Gyakorlati alkalmazások

Az evolúciós algoritmusokat széles körben, az üzleti szférától az autógyártásig alkalmazzák. Immáron egyre több befektető jobban megbízik a szimulált evolúció "túlélőiben", mint a pénzügyi szakértőkben. A 3,7 trillió dollár tőkével rendelkező amerikai State Street Global Advisors cég vezetői a beszerzésre és az értékesítésre vonatkozó döntések meghozatalakor mind gyakrabban támaszkodnak evolúciós algoritmusokra. A Barclays Global Investors, a Fidelity és a PanAgora Asset Management úgyszintén.

A General Motorsnál az autók festésének koordinációját végzik evolúciós algoritmusok segítségével, s az esetleges színcserék (magas) költségeit mintegy ötven százalékkal csökkentik így.

A Volvo a 770 teherautó-típus gyártásánál, a hárombillió dollár tőkeerejű cement-társaság, a Cemex - a szállítás logisztikai elemeinek a kidolgozásánál - szintén evolúciós algoritmusokat használ fel.

A DNS egy genetikai program, egy, a mai emberi kapacitással utánozhatatlan, utolérhetetlen komplexitású software.

A biológiai modellen túllépve, az eddigi kísérletek eredményei alapján megállapítható, hogy az evolúciós minta némely számítógépes területen / környezetben (gépi kódok, stb.) sikerrel alkalmazható. Szinte mindig "megtalálja" a

kellő formákat és folyamatokat, s teszi ezt úgy, hogy közben maximálisan ki is használja a médium nyújtotta lehetőségeket.

A rendszertervezők egyetlen megoldást se programoznak be: a megoldások a (kíméletlen) verseny és a (szüntelen) bizonyítás során fejlődnek ki. Ne ijedjünk meg: a klasszikus modell évmillióihoz viszonyítva, felgyorsított folyamatról van szó. Több ezer generáció szimulálására - a hardware-kapacitás függvényeként - órák, napok, esetleg hetek alatt kerül sor. És a bonyolult iterációs mechanizmusokon is mindössze egyszer kell végigmennünk, utána már "csak" a kifejlődött, kifinomult szabályokat alkalmazzuk.

Önszerveződő, "valahonnan" felbukkanó módszereket... Hiszen az eredményt lehetetlen előre jelezni, sokszor még a tervezők számára is teljesen meglepő. Akárcsak a processzus, melyen a "diadalmas" programok keresztülmennek. Gyakran fordul elő, hogy már száz-kétszáz iteráción vagyunk túl, és látszólag szinte semmi nem történt. Aztán hirtelen, egyik pillanatról a másikra, kikristályosodik a megoldás.

A Tierra Hálózat

Thomas Ray, amerikai biológus és a trópusi esőerdők szakértője néhány évvel ezelőtt merész vállalkozásba, a Tierra Hálózat "kiépítésébe", azaz az evolúció evolúciójának a szimulálásába fogott. Ebben a természetes organizmusok digitális megfelelői, azok számítógépes szimulációi által belakott, több mint százötven, a világhálóra csatlakozott computer alkotta cybertérben minden egyes "sejt" DNS-szerű genetikai kóddal rendelkezik.

A lények egyrészt a mesterséges és kaotikus környezet limitált tereiért, másrészt a szintén limitált energiakészletekért küzdenek meg / versengenek egymással. Közben a legprimitívebb egysejtű formákból változatos, többsejtű, több funkciót végrehajtó szervezetekké alakulnak.

A projekt egyik célja, hogy a szimuláció segítségével a veszélyeztetett esőerdők megóvására életképes modelleket dolgozzunk ki. A másik, hogy jobban megértsük a hozzávetőleg hétszázötvenmillió évvel ezelőtti, Kambrium-kori eseményeket, az evolúció robbanásszerű "begyorsulását": a többsejtű állatok megjelenését, a különböző és rendkívül változatos életformák hirtelen, előzmények nélküli elszaporodását.

¤¤

Kurzweil's Turing Fallacy

http://life.ou.edu/pubs/kurzweil/

Thomas Ray
Department of Zoology
University of Oklahoma
Norman, Oklahoma 73019
tray@ou.edu

https://en.wikipedia.org/wiki/Thomas_S._Ray
http://life.ou.edu/

There are numerous directions from which to criticize Kurzweil's book. In this essay I will focus on his failure to consider the unique nature of the digital medium when discussing artificial intelligence. But before elaborating on this point, I would like to briefly call attention to some other issues.

Psychic Quantum Mechanics

Kurzweil's interpretation of quantum mechanics leads him to the conclusion that "consciousness, matter, and energy are inextricably linked". While this is true in the sense that consciousness arises from the interactions of matter and energy, it is not true in the sense that Kurzweil intends it: that quantum ambiguities are not resolved until they are forced to do so by a conscious observer.

Kurzweil's error is most glaringly apparent in his description of the paper output from a quantum computer: "So the page with the answer is ambiguous, undetermined -

until and unless a conscious entity looks at it. Then instantly all the ambiguity is retroactively resolved, and the answer is there on the page. The implication is that the answer is not there until we look at it." He makes the same error in describing the evolution of the universe: "from one perspective of quantum mechanics - we could say that any Universe that fails to evolve conscious life to apprehend its existence never existed in the first place."

Kurzweil does not understand that it is the act of measurement that causes the collapse of the wave function, not conscious observation of the measurement. In practice, the collapse is (probably always) caused by a completely unconscious measuring device. Printing of the result on a paper could be such a measuring device. Subsequent conscious observation of the measurement is irrelevant.

This psychic quantum mechanics did not originate with Kurzweil. It has been around for decades, apparently as a way to deal with Schrodinger's cat. Thus, Kurzweil may be able to point to physicists who hold this view. Similarly, I could point to biologists who believe in the biblical story of creation rather than evolution.

Colloquial Chaos

Kurzweil's suggestion that in a process, the time interval between salient events expands or contracts along with the amount of chaos ("the law of time and chaos"), is quite interesting. Yet, the definitions of "salient events" and "chaos" are quite subjective, making the "law" difficult to support. Technically, it would probably be more appropriate to use the word "entropy" in place of "chaos", but for consistency, I will also use "chaos" in this discussion.

Most striking is the apparently inconsistent use of chaos. He states that in an evolutionary process order increases, and he says: "evolution draws upon the chaos in the larger system in which it takes place for its options for diversity". Yet he states that in the development of an individual organism chaos increases, and he says: "The development of an organism from conception as a single cell through maturation is a process moving toward greater diversity and thus greater disorder." Kurzweil suggests that in evolution, diversity implies order, while in development, diversity implies disorder.

Through evolution, the diversity of species on Earth has increased, and through development, the diversity of cell types increases. I would characterize both as processes that generate order. Why does Kurzweil consider development to generate

chaos? Apparently in order to make his law of time and chaos consistent with our perception of time: our subjective unit of time grows with our age.

I believe that the scientific community would be in general agreement that the developmental process up to the period of reproduction is a process of increasing order. In humans, who live well beyond the reproductive years, the condition of the body begins to deteriorate after the reproductive years, and this senescence would generally be considered a process of increasing chaos.

In an effort to fit development seamlessly into his law of time and chaos, Kurzweil presents the whole life cycle from conception to death, as unidirectional, towards increasing chaos. This position is indefensible. The developmental process directly contradicts the law of time and chaos. Development is a process in which the time between salient events increases with order.

He attempts to be clear and concrete in his use of the term chaos: "If we're dealing with the process of evolution of life-forms, then chaos represents the unpredictable events encountered by organisms, and the random mutations that are introduced in the genetic code." He explains: "Evolution draws upon the great chaos in its midst - the ever increasing entropy governed by the flip side of the Law of Time and Chaos - for its options for innovation." This implies that unpredictable events and mutations are becoming more frequent, a position that would be difficult to defend. His argument is that increasing rates of mutations and unpredictable events are, in part, driving the increasing frequency of "salient events" in evolution. He does not provide any support for this highly questionable argument.

Despite his attempt to be precise, his use of "chaos" is vernacular: "When the entire Universe was just a 'naked' singularity ... there was no chaos." "As the Universe grew in size, chaos increased exponentially." "Now with billions of galaxies sprawled out over trillions of light-years of space, the Universe contains vast reaches of chaos..." "We start out as a single fertilized cell, so there's only rather limited chaos there. Ending up with trillions of cells, chaos greatly expands." It seems that he associates chaos with size, a very unconventional use of the term.

His completely false interpretation of quantum mechanics, his vague and inconsistent use of terms such as "chaos" and "salient events", and his failure to understand the thermodynamics of development represent errors in the basic science from which he constructs his view of the world. These misunderstandings of basic science seriously undermine the credibility of his arguments.

I am not comfortable with the equation of technological development and evolution. I think that most evolutionary biologists would consider these to be quite separate processes, yet, their equation represents a point of view consistent with Kurzweil's arguments, and also consistent with the concept of "meme" developed by the evolutionary biologist Richard Dawkins.

Intelligence in the Digital Medium

The primary criticism that I wish to make of Kurzweil's book, is that he proposes to create intelligent machines by copying human brains into computers. We might call this the Turing Fallacy. The Turing Test suggests that we can know that machines have become intelligent when we can not distinguish them from human, in free conversation over a teletype. The Turing Test is one of the biggest red-herrings in science.

It reminds me of early cinema when we set a camera in front of a stage and filmed a play. Because the cinema medium was new, we completely didn't understand what it is and what we can do with it. At that point we completely didn't understand the nature of the medium of cinema. We are almost in the same position today with respect to the digital medium.

Over and over again, in a variety of ways, we are shaping cyberspace in the form of the 3D material space that we inhabit. But cyberspace is not a material space and it is not inherently 3D. The idea of downloading the human mind into a computer is yet another example of failing to understand and work with the properties of the medium. But let me give some other examples and then come back to this.

I have heard it said that cyberspace is a place for the mind, yet we feel compelled to take our bodies with us. 3D virtual worlds and avatars are manifestations of this. I have seen virtual worlds where you walk down streets lined by buildings. In one I saw a Tower Records store, whose front looked like the real thing. You approached the door, opened it, entered, and saw rows of CDs on racks, and an escalator to take you to the next floor. Just Like The Real Thing!

I saw a demo of Alpha World, built by hundreds of thousands of mostly teenagers. It was the day after Princess Diana died, and there were many memorials to her, bouquets of flowers by fountains, photos of Diana with messages. It looked Just Like The Real memorials to Diana.

I wondered, why do these worlds look and function as much as possible like the real thing? This is cyberspace, where we can do anything. We can move from point A to point B instantly without passing through the space in between. So why are we forcing ourselves to walk down streets and halls and to open doors?

Cyberspace is not a 3D Euclidean space. It is not a material world. We are not constrained by the same laws of physics, unless we impose them upon ourselves. We need to liberate our minds from what we are familiar with before we can use the full potential of cyberspace. Why should we compute collision avoidance for avatars in virtual worlds when we have the alternative to find out how many avatars can dance on the head of a pin?

The WWW is a good counter-example, because it recognizes that in cyberspace it doesn't matter where something is physically located. Amazon.com is a good alternative to the mindlessly familiar 3D Tower Record store.

Let me come back to Kurzweil's ideas on AI. Kurzweil states that it is "ultimately feasible" to: "...scan someone's brain to map the locations, interconnections, and contents of the somas, axons, dendrites, presynaptic vesicles, and other neural components. Its entire organization could then be re-created on a neural computer of sufficient capacity, including the contents of its memory.... we need only to literally copy it, connection by connection, synapse by synapse, neurotransmitter by neurotransmitter."

This passage most clearly illustrates Kurzweil's version of the Turing Fallacy. It is not only infeasible to "copy" a complex organic organ into silicon without losing its function, but it is the least imaginative approach to creating an AI. How do we copy a serotonin molecule or a presynaptic vesicle into silicon? This passage of the book does not explicitly state whether he is proposing a software simulation from the molecular level up, of a copy of the brain, or if he is proposing the construction of actual silicon neurons, vesicles, neurotransmitters, and their wiring together into an exact copy of a particular brain. Yet in the context of the preceding discussion, it appears that he is proposing the latter.

Such a proposal is doomed to failure. It would be a fantastic task to map the entire physical, chemical, and dynamic structure of a brain. Even if this could be accomplished, there would be no method for building a copy. There is no known technology for building complexly differentiated microscopic structure on such a

large scale. If a re-construction method existed, we might expect that a copy made of the same materials, carbon chemistry, if somehow jump-started into the proper dynamic activity, would have the same function (though such a copied brain would require a body to support it). But a copy made of metallic materials could not possibly have the same function. It would be a fantastically complex and intricate dynamic sculpture, whose function would bear no relation to a human brain. And what of the body and its essential sensory integration with the brain?

In order for the metallic "copy" to have the same function, we would have to abstract the functional properties out of the organic neural elements, and find structures and processes in the new metallic medium that provide identical functions. This abstraction and functional-structural translation from the organic into the metallic medium would require a deep understanding of the natural neural processes, combined with the invention of many computing devices and processes which do not yet exist.

However, Kurzweil has stated that one advantage of the brain-copy approach is that "we don't need to understand all of it; we need only to literally copy it..." Yet he is ambivalent on this critical point, adding: "To do this right, we do need to understand what the salient information-processing mechanisms are. Much of a neuron's elaborate structure exists to support its own structural integrity and life processes and does not directly contribute to its handling of information."

The structure and function of the brain or its components can not be separated. The circulatory system provides life support for the brain, but it also delivers hormones which are an integral part of the chemical information processing function of the brain. The membrane of a neuron is a structural feature defining the limits and integrity of a neuron, but it is also the surface along which depolarization propagates signals. The structural and life-support functions can not be separated from the handling of information.

The brain is a chemical organ, with a broad spectrum of chemical communication mechanisms ranging from microscopic packets of neurotransmitters precisely delivered at target synapses, to NO gas and hormones spread through the circulatory system or diffusing through the intercellular medium of the brain. There also exist a wide range of chemical communications systems with intermediate degrees of specificity of delivery. The brain has evolved its exquisitely subtle and complex functionality based on the properties of these chemical systems. A metallic computation system operates on fundamentally different dynamic properties, and

could never precisely and exactly "copy" the function of a brain.

The materials of which computers are constructed have fundamentally different physical, chemical, and electrical properties than the materials from which the brain is constructed. It is impossible to create a "copy" of an organic brain out of the materials of computation. This applies not only to the proposition of copying an individual human brain with such accuracy as to replicate a human mind along with its memories, but also to the somewhat less extreme proposition of creating an artificial intelligence by reverse engineering the human brain.

Structures and processes suitable for information processing in the organic medium are fundamentally different from those of the metallic computational medium. Intelligent information processing in the computational medium must be based on fundamentally different structures and processes, and thus can not be copied from organic brains.

I see three separate processes which are sometimes confounded. Machines having:

computing power equal to the level of human intelligence
computing performance equal to the level of human intelligence
computing like human intelligence
A large portion of Kurzweil's book establishes the first process by extrapolating Moore's law into the future until individual machines can perform the same number of computations per second as is estimated for the human brain (~2020 AD).

I accept that this level of computing power is likely to be reached, someday. But no amount of raw computer power will be intelligent unless it is properly organized. This is a software problem, not a hardware problem. The organizational complexity of software does not march forward according to Moore's law.

While I can accept that computing power will inevitably reach human levels, I am not confident that computing performance will certainly follow. The exponential increase of computing power is driven by higher densities and greater numbers of components on chips, not by exponentially more complex chip designs.

The most complex of artifacts designed and built by humans are much less complex that living organisms. Yet the most complex of our creations are showing alarming failure rates. Orbiting satellites and telescopes, space shuttles, interplanetary probes, the pentium chip, computer operating systems, all seem to be pushing the limits of

what we can effectively design and build through conventional approaches.

It is not certain that our most complex artifacts will be able to increase in complexity by an additional one, two, or more orders of magnitude, in pace with computing power. Our most complex software (operating systems and telecommunications control systems) already contain tens of millions of lines of code. At present it seems unlikely that we can produce and manage software with hundreds of millions or billions of lines of code. In fact there is no evidence that we will ever be able to design and build intelligent software.

This leads to the next distinction, which is central to my message, and requires some explanation:

computing performance equal to the level of human intelligence
computing like human intelligence
A machine might exhibit an intelligence exactly like and indistinguishable from humans, a Turing AI, or a machine might exhibit a fundamentally different kind of intelligence, like some science fiction alien intelligence. I expect that intelligences which emerge from the digital and organic media will be as different as their respective media, even if they have comparable computing performance.

Everything we know about life is based on one example of life: Life on Earth. Everything we know about intelligence is based on one example of intelligence: Human intelligence. This limited experience burdens us with preconceptions and limits our imaginations.

Consider this thought experiment:

We are all robots. Our bodies are made of metal and our brains of silicon chips. We have no experience or knowledge of carbon based life, not even in our science fiction. Now one of us robots comes to an AI discussion with a flask of methane, ammonia, hydrogen, water, and some dissolved minerals. The robot asks: "Do you suppose we could build a computer from this stuff?"

The engineers among us might propose nano-molecular devices with fullerene switches, or even DNA-like computers. But I am sure they would never think of neurons. Neurons are astronomically large structures compared to the molecules we are starting with.

Faced with the raw medium of carbon chemistry, and no knowledge of organic life, we would never think of brains built of neurons, supported by circulatory and digestive systems, in bodies with limbs for mobility, bodies which can only exist in the context of the ecological community that feeds them.

We are in a similar position today as we face the raw medium of digital computation and communications. The preconceptions and limited imagination deriving from our organic-only experience of life and intelligence make it difficult for us to understand the nature of this new medium, and the forms of life and intelligence that might inhabit it.

How can we go beyond our conceptual limits, find the natural form of intelligent processes in the digital medium, and work with the medium to bring it to its full capacity, rather than just imposing the world we know upon it by forcing it to run a simulation of our physics, chemistry, and biology?

In the carbon medium it was evolution that explored the possibilities inherent in the medium, and created the human mind. Evolution listens to the technology that it is embedded in. It has the advantage of being mindless, and therefore devoid of preconceptions, and not limited by imagination.

I propose the creation of a digital nature. A system of wildlife reserves in cyberspace, in the interstices between human colonizations, feeding off of unused cpu-cycles (and permitted a share of our bandwidth). This would be a place where evolution can spontaneously generate complex information processes, free of the demands of human engineers and market analysts telling it what the target applications are.

Digital naturalists can then explore this cyber-nature in search of applications for the products of digital evolution in the same way that our ancestors found applications among the products of organic nature such as: rice, wheat, corn, chickens, cows, pharmaceuticals, silk, mahogany. But, of course, the applications that we might find in the living digital world would not be material, they would be information processes.

It is possible that out of this digital nature there might emerge a digital intelligence, truly rooted in the nature of the medium, rather than brutishly copied and downloaded from organic nature. It would be a fundamentally alien intelligence, but one which would complement rather than duplicate our talents and abilities.

I think it would be fair to say that the main point of Kurzweil's book is that artificial entities with intelligence equal to and greater than humans will inevitably arise, in the near future. While his detailed explanation of how this might happen focuses on what I consider to be the Turing Fallacy that it will initially take a human form, Kurzweil would probably be content with any route to these higher intelligences, Turing or non-Turing.

While I feel that AIs must certainly be non-Turing (not like human intelligences), I feel ambivalent about whether they will emerge at all. It is not the certainty that Kurzweil paints, like the inexorable march of Moore's law. Raw computing power is not intelligence. Our ability to ever create information processes of a complexity comparable to the human mind is completely unproven and absolutely uncertain.

I have suggested evolution as an alternate approach to producing intelligent information processes. These evolved AIs would certainly be non-Turing AIs. Yet evolution in the digital medium remains a process with a very limited record of accomplishments. We have been able to establish active evolutionary processes, by both natural and artificial selection in the digital medium. But the evolving entities have always contained at most several thousand bits of genetic information.

We do not yet have a measure on the potential of evolution in this medium. If we were to realize a potential within several orders of magnitude of that of organic evolution, it would be a spectacular success. But if the potential of digital evolution falls ten orders of magnitude below organic evolution, then digital evolution will lose its luster. There is as yet no evidence to suggest which outcome is more likely.

The hope for evolution as a route to AI is not only that it would produce an intelligence rooted in and natural to the medium, but that evolution in the digital medium is capable of generating levels of complexity comparable to what it has produced in the organic medium. Evolution is the only process that is proven to be able to generate such levels of complexity. However, that proof is in the organic, not the digital medium. Like an artist that can express their creativity in oil paint but not stone sculpture, evolution may be capable of magnificent creations in the organic medium but not the digital.

Yet the vision of digital evolution of vast complexity is still out there, waiting for realization or disproof. We are only at the most rudimentary level of our experience with evolution in the digital medium. The possibilities are great enough to merit a serious and sustained effort.

¤¤¤

ARE WE SPIRITUAL MACHINES? | Chapter 9: What Turing Fallacy?: Response to Thomas Ray
June 7, 2001
http://www.kurzweilai.net/chapter-9-what-turing-fallacy-response-to-thomas-ray

author | Ray Kurzweil
year published | 2001
Measurement Without Observation

In Thomas Ray's articulated world, there is no such thing as consciousness, a view he makes clear in his reductionist view of quantum mechanics. Ray states that "it is the act of measurement that causes the collapse of the wave function, not conscious observation of the measurement." In other words, according to Ray, the collapse of the wave function is caused by measurement without observation, but what could this refer to? We know that any quantum "collapsed" event involving one or more particles causes some reaction beyond those particles immediately involved. No particle is an "island," so to speak. These inevitable reactions are properly considered measurements. The only conceivable circumstance in which an event would not cause a specific reaction (i.e., a measurement) is if the event was indeterminate because the wave function was not collapsed. It is only through the collapse of the wave function that an event becomes determinate and thereby causes a reaction in the world, which constitutes a measurement. It is, therefore, the collapse of the wave function that causes measurement, which Ray tells us causes collapse of the wave function. So what Ray is saying is that the collapse of the wave function causes the collapse of the wave function. By removing the concept of observation from measurement, Ray's explanation of quantum mechanics devolves into this tautology.

Ray goes on to call any other view, even those of other mainstream scientists, a "glaring…error." Ray's rigid view regards any introduction of consciousness in the world to be an "error." I would also point out that if we accept Ray's view, then Penrose's objection to the potential of consciousness in a nonbiological entity (i.e., Penrose's argument to the effect that such an entity would have to recreate the precise

quantum state of a biological conscious entity) becomes even less valid.

Colloquial Chaos Indeed

Ray casts doubt that there is increased chaos as an organism moves from conception as a single cell to a mature individual. Consider what we know about this process. The human genome contains 3 billion DNA rungs for a total of 6 billion bits of data. There is enormous redundancy in this information (e.g., a sequence known as "ALU" is repeated hundreds of thousands of times), so the amount of unique information is estimated at around 3% or about 23 megabytes. In contrast, the human brain contains on the order of 100 trillion connections. Just specifying this connection data would require trillions of bytes. Thus as we go from the genome, which specifies the brain among all other organs, to the fully expressed individual, the amount of information, considering just the brain connection patterns alone, increases by a factor of millions. We know that the genome specifies a wiring plan for the interneuronal connections that includes a great deal of randomness, i.e., chaos, at specific implementation stages. This includes the stage of fetal wiring, during which interneuronal connections essentially wire themselves with a significant element of randomness applied during the process, as well as the growing of new dendrites after birth (which is believed to be a critical part of the learning process). This is at least one source of the increasing chaos resulting from the development of the individual from a fertilized egg. Another source is the chaos inherent in the environment that the individual encounters.

Ray states that I argue that "increasing rates of mutations and unpredictable events are, in part, driving the increasing frequency of 'salient events' in evolution." That's not my argument at all and I never make this statement. My position is that the acceleration of an evolutionary process, including both biological and technological evolution, results from the greater power of each level of evolution to create the next level. For example, with the evolution of DNA, evolutionary experiments could proceed more rapidly and more effectively with each stage of results recorded in the evolving DNA code. With the innovation of sexual reproduction, a more effective means for devising new combinations of genetic information became available. Within technological evolution, we also find that each generation of technology enables the next generation to proceed more rapidly. For example, the first generation of computers were designed by pen on paper and built with screwdrivers and wires. Compare that to the very rapid design of new computers using today's computer-assisted design tools.

Ray's Digital Fallacy

Thomas Ray starts his chapter by citing my alleged "failure to consider the unique nature of the digital medium." However, my thesis repeatedly refers to combining analog and digital methods in the same way that the human brain does; for example, "more advanced neural nets [which] are already using highly detailed models of human neurons, including detailed nonlinear analog activation functions." I go on to cite the advantages of emulating the brain's "digital controlled analog" design, and conclude that "there is a significant efficiency advantage to emulating the brain's analog methods." Analog methods are not the exclusive province of biological systems. We used to refer to "digital computers" to distinguish them from the more ubiquitous analog computers which were widely used during World War II.

It is also worth pointing out that analog processes can be emulated with digital methods whereas the reverse is not necessarily the case. However, there are efficiency advantages to analog processing as I point out above. Analog methods are readily recreated by conventional transistors which are essentially analog devices. It is only by adding the additional mechanism of comparing the transistor's output to a threshold that it is made into a digital device.

What Turing Fallacy?

Thomas Ray states:

The primary criticism that I wish to make of Kurzweil's book, is that he proposes to create intelligent machines by copying human brains into computers. We might call this the Turing Fallacy. The Turing Test suggests that we can know that machines have become intelligent when we cannot distinguish them from human, in free conversation over a teletype. The Turing Test is one of the biggest red-herrings in science.

This paragraph contains several ideas, unrelated ones in my view, but concepts that appear to animate Ray's discomfort with "strong AI." I should make the caveat that Ray appears to accept the possibility of an advanced but "fundamentally alien intelligence" that is "rooted in and natural to the medium," but he dismisses the possibility, as well as the desirability, of AIs sharing human-like attributes. He states that "AIs must certainly be non-Turing," which he defines as "unlike human intelligences." So this is what Ray means by the "Turing Fallacy." He is maintaining that any intelligence that might emerge in nonbiological mediums would not and

could not be a human intelligence, and would, therefore, be unable to pass the Turing Test.

It is fair to conclude from this that Thomas Ray accepts the Turing Test as a reasonable test for "human intelligence," but makes the point that there could be an alien intelligence that is very capable in terms of performing intelligent tasks but unable to pass this particular test. I should point out that Turing has made precisely the same point. Turing intended his Test specifically as a measure of human intelligence. An entity passing the Turing Test may be said to be intelligent, and moreover, to possess a human-form of intelligence. Turing specifically states that the converse statement does not hold, that failure to pass the Test does not indicate a lack of intelligence. I've made the same point. As I stated in the Age of Intelligent Machines, certain animals such as dolphins, giant squids, and certain species of whales, appear to have relatively high levels of intelligence, but are in no position to pass the Turing Test (they can't type for one thing). Even a human would not be able to pass the Test if she didn't speak the language of the Turing Test judge. The key point of the Turing test is that human language is sufficiently broad that we can test for the full range of human intelligence through human language dialogues. The Turing Test itself does not represent a fallacy, but rather a keen insight into the power of human communication through language to represent our thinking processes.

So where is the fallacy? Ray appears to be objecting to the concept of creating a human-like intelligence, one whose communication is sufficiently indistinguishable from human language-based communication such that it could pass the Turing Test, on grounds of both desirability and feasibility.

With regard to the first issue, the desirability of machines understanding our language is, I believe, clear from the entire history of the AI field. As machines have gained proficiency in aspects of human language, they have become more useful and more valuable. Language is our primary method of communication, and machines need to understand human language in order to interact with us in an efficacious manner. Ultimately, they will need to understand human knowledge in a deep way to fully manifest their potential to be our partners in the creation of new knowledge. And as Turing pointed out, understanding human knowledge, including our capacity for understanding and expressing higher order emotions, is a prerequisite for effective human communication, and is therefore, a necessary set of skills to pass the Turing Test.

With regard to the issue of feasibility, Thomas Ray states:

I accept that this level of computing power is likely to be reached, someday. But no amount of raw computer power will be intelligent in the relevant sense unless it is properly organized. This is a software problem, not a hardware problem.

I agree with this, of course, as I've had to state several times. A primary scenario that I describe in solving the "software problem" is to reverse engineer the methods deployed by the human brain. Although I also talk about the potential to copy specific human brains, I acknowledge that this is a more difficult task and will take longer. Reverse engineering the human brain is quite feasible, and we are further along in this project than most people realize. As I pointed out earlier, there are many contemporary examples that have demonstrated the feasibility of reverse engineering human neurons, neuron clusters, and entire brain regions, and then implementing the resulting detailed mathematical models in nonbiological mediums.

I mentioned Lloyd Watts' work in which he has developed a detailed and working model of more than a dozen brain regions related to auditory processing. Carver Mead's retina models, which are implemented as digital controlled analog processes on silicon chips, capture processes similar to those that take place in human visual processing. The complexity and level of detail in these models is expanding exponentially along with the growing capacity of our computational and communication technologies. This undertaking is similar to the genome project in which we scanned the genome and are now proceeding to understand the three-dimensional structures and processes described therein. In the mission to scan and reverse engineer the neural organization and information processing of the human brain, we are approximately where we were in the genome project about ten years ago. I estimate that we will complete this project within thirty years, which takes into account an exponential rather than linear projection of our anticipated progress.

Thomas Ray essentially anticipates the answer to his own challenge, when he states that:

In order for the metallic "copy" to have the same function, we would have to abstract the functional properties out of the organic neural elements, and find structures and processes in the new metallic medium that provide identical functions. This abstraction and functional-structural translation from the organic into the metallic medium would require a deep understanding of the natural neural processes, combined with the invention of many computing devices and processes which do not yet exist.

I'm not sure why Ray uses the word "metallic" repeatedly, other than to demonstrate his inclination to regard nonbiological intelligence as inherently exhibiting the brittle, mechanical, and unsubtle properties that we have traditionally associated with machines. However, in this paragraph, Ray describes the essential process that I have proposed. Many contemporary projects have shown the feasibility of developing and expressing a deep understanding of "natural neural processes," and the language for that expression is mathematics.

We clearly will need to invent new computing devices to create the necessary capacity, and Ray appears to accept that this will happen. As for inventing new processes, new algorithms are continually being developed as well; however, what we have found thus far is that we have encountered few difficulties instantiating these models once they are revealed. The mathematical models derived from the reverse engineering process are readily implemented using available methods.

The revealed secrets of human intelligence will undoubtedly provide many enabling methods in the creation of the software of intelligence. An added bonus will be deep insight into our own nature, into human function and dysfunction.

Thomas Ray states that:

The structure and function of the brain or its components cannot be separated. The circulatory system provides life support for the brain, but it also delivers hormones that are an integral part of the chemical information processing function of the brain. The membrane of a neuron is a structural feature defining the limits and integrity of a neuron, but it is also the surface along which depolarization propagates signals. The structural and life-support functions cannot be separated from the handling of information.

Ray goes on to describe several of the "broad spectrum of chemical communication mechanisms" that the brain exhibits. However, all of these are readily modelable, and a great deal of progress has already been made in this endeavor. The intermediate language is mathematics, and translating the mathematical models into equivalent nonbiological mechanisms is the easiest step in this process. With regard to the delivery of hormones by the circulatory system, this is an extremely low bandwidth phenomenon, which will not be difficult to model and replicate. The blood levels of specific hormones and other chemicals influence parameter levels that affect a great many synapses at once.

Ray concludes that "a metallic computation system operates on fundamentally different dynamic properties and could never precisely and exactly 'copy' the function of a brain." If one follows closely the progress in the related fields of neurobiology, brain scanning, neuron and neural region modeling, neuron-electronic communication, neural implants, and related endeavors, we find that our ability to replicate the salient functionality of biological information processing can meet any desired level of precision. In other words, the copied functionality can be "close enough" for any conceivable purpose or goal, including satisfying a Turing Test judge. Moreover, we find that efficient implementations of the mathematical models require substantially less computational capacity than the theoretical potential of the biological neuron clusters being modeled.

Ray goes on to describe his own creative proposal for creating nonbiological intelligence, which is to use evolutionary algorithms to allow a digital intelligence to "emerge," one that is "rooted in the nature of the medium." This would be, according to Ray, a "non-Turing intelligence," but "one which would complement rather than duplicate our talents and abilities."

I have no problem with this particular idea. Another of Ray's mistaken claims is that I offer human brain reverse engineering as the only route to strong AI. The truth is that I strongly advocate multiple approaches. I describe the reverse engineering idea (among others) because it serves as a useful existence proof of the feasibility of understanding and replicating human intelligence. In my book The Age of Spiritual Machines, I describe a number of other approaches, including the one that Ray prefers (evolutionary algorithms). My own work in pattern recognition and other aspects of AI consistently utilizes multiple approaches, and it is inevitable that the ultimate path to strong AI will combine insights from a variety of paradigms. The primary role of developing mathematical models of biological neurons, scanning the human brain, and reverse engineering the hundreds of brain regions is to develop biologically-inspired models of intelligence, the insights of which we will then combine with other lines of attack.

With regard to Ray's own preference for evolutionary or genetic algorithms, he misstates the scope of the problem. He suggests that the thousands of bits of genetic information in contemporary genetic algorithms may end up falling "ten orders of magnitude below organic evolution." But the thousands of bits of genetic order represented by contemporary systems are already only four or five orders of magnitude below that of the unique information contained in the genome.

Ray makes several other curious statements. Ray says:

…the most complex of our creations are showing alarming failure rates. Orbiting satellites and telescopes, space shuttles, interplanetary probes, the Pentium chip, computer operating systems, all seem to be pushing the limits of what we can effectively design and build through conventional approaches… Our most complex software (operating systems and telecommunications control systems) already contains tens of millions of lines of code. At present it seems unlikely that we can produce and manage software with hundreds of millions or billions of lines of code.

First of all, what alarming failure rates is Ray referring to? Computerized mission critical systems are remarkably reliable. Computerized systems of significant sophistication routinely fly and land our airplanes automatically. I am not aware of any airplane crash that has been attributed to a failure of these systems, yet many crashes are caused by the human errors of pilots and maintenance crews. Automated Intensive Care monitoring systems in hospitals almost never malfunction, yet hundreds of thousands of people die from human medical errors. If there are alarming failure rates to worry about, it's human failure, not those of mission critical computer systems. The Pentium chip problem which Ray alludes to was extremely subtle, caused almost no repercussions, and was quickly rectified.

The complexity of computerized systems has indeed been scaling up exponentially. Moreover, the cutting edge of our efforts to emulate human intelligence will utilize the self-organizing paradigms that we find in the human brain. I am not suggesting that self-organizing methods such as neural nets and evolutionary algorithms are simple or automatic to use, but they represent powerful tools which will help to alleviate the need for unmanageable levels of complexity.

Most importantly, it is not the case that the human brain represents a complexity comparable to "billions of lines of code." The human brain is created from a genome of only about 23 million bytes of unique information (less than Microsoft Word). It is through self-organizing processes that incorporate significant elements of randomness (as well as exposure to the real world) that this small amount of design information is expanded to the trillions of bytes of information represented in a mature human brain. The design of the human brain is not based on billions of lines of code or the equivalent thereof. Similarly, the task of creating human level intelligence in a nonbiological entity will not involve creating a massive expert system comprising billions of rules or lines of code, but rather a learning, chaotic,

self-organizing, system, one ultimately that is biologically inspired.

Thomas Ray writes:

The engineers among us might propose nano-molecular devices with fullerene switches, or even DNA-like computers. But I am sure they would never think of neurons. Neurons are astronomically large structures compared to the molecules we are starting with.

This is exactly my own point. The purpose of reverse engineering the human brain is not to copy the digestive or other unwieldy processes of biological neurons, but rather to understand their salient information processing methods. The feasibility of doing this has already been demonstrated in dozens of contemporary projects. The scale and complexity of the neuron clusters being emulated is scaling up by orders of magnitude, along with all of our other technological capabilities.

Once we have completed the reverse engineering of the several hundred brain regions, we will implement these methods in nonbiological substrates. In this way, we will combine these human-like capacities with the natural advantages of machines, i.e., the speed, accuracy and scale of memory, and, most importantly, the ability to instantly share knowledge.

Ray writes:

Over and over again, in a variety of ways, we are shaping cyberspace in the form of the 3D material space that we inhabit. But cyberspace is not a material space and it is not inherently 3D. The idea of downloading the human mind into a computer is yet another example of failing to understand and work with the properties of the medium….Cyberspace is not a 3D Euclidean space. It is not a material world. We are not constrained by the same laws of physics, unless we impose them upon ourselves.

The reality is that we can do both. At times we will want to impose upon our virtual reality environments the three-dimensional gravity-bound reality we are used to. After all, that is the nature of the world we are comfortable in. At other times, we may wish to explore environments that have no earthly counterpart, ones that may indeed violate the laws of physics.

We can also do both with regard to emulating intelligence in our machines. We can apply Ray's preferred genetic algorithm approach while we also benefit from the

reverse engineering of biological information processes, among other methods.

In summing up, Thomas Ray writes:

Everything we know about intelligence is based on one example of intelligence, namely, human intelligence. This limited experience burdens us with preconceptions and limits our imaginations.

Actually, it is Thomas Ray who is limiting his imagination to his single idea of unleashing "evolution in the digital medium." Certainly there will be new forms of intelligence as nonbiological intelligence continues to grow. It will draw upon many sources of knowledge, some biologically motivated, and some inspired by our own imagination and creativity, ultimately augmented by the creativity of our machines.

The power of our civilization has already been greatly augmented by our technology, with which we are becoming increasingly intimate, as devices slip into our pockets and dangle from our bodies like jewelry. Within this decade, computing and communications will appear to disappear, with electronics being woven into our clothing, images being written directly to our retinas, and extremely high bandwidth wireless communication continually augmenting our visual and auditory reality. Within a few decades, with technology that travels inside our bodies and brains, we will be in a position to vastly expand our intelligence, our experiences, and our capacity to experience. Nonbiological intelligence will ultimately combine the inherent speed and knowledge sharing advantages of machines with the deep and subtle powers of the massively parallel, pattern recognition-based, biological paradigm.

Today, our most sophisticated machines are still millions of times simpler than human intelligence. Similarly, the total capacity of all the computers in the world today remains at least six orders of magnitude below that of all human brains, which I estimate at 1026 digitally controlled analog transactions per second. However, our biological capacity is fixed. Each of us is limited to a mere hundred trillion interneuronal connections. Machine intelligence, on the other hand, is growing exponentially in capacity, complexity, and capability. By the middle of this century, it will be nonbiological intelligence, representing an intimate panoply of paradigms, that predominates.