Premi UPC de ciència-ficció 2002

Conferència, novembre 2002

Vernor Vinge
La singularitat tecnològica

[Aquest treball es basa en la meva presentació el 1993 al Simposi VISION-21 patrocinat pel NASA Lewis Research Center i l'Institut Aerospacial d'Ohio, els dies 30 i 31 de març de 1993 [N7]. Els principals canvis que he introduït en aquesta presentació per a la UPC tenen la finalitat d'aconseguir que el meu discurs es correspongui amb les imatges de la meva presentació i d'oferir alguns exemples nous i una bibliografia ampliada.]

La llei de Moore [N1] diu en línies generals que el nombre de microxips d'un circuit integrat es duplicarà cada dos anys. Actualment, la frase "la llei de Moore" sovint s'aplica en afirmacions sobre el creixement exponencial de la potència del hardware dels ordinadors i dels equips de comunicacions. Un creixement com aquest té implicacions dràstiques, però sempre és savi mostrar prudència en els raonaments sobre les corbes de tendències. Aquests raonaments no tenen res semblant a la fiabilitat de les prediccions en física i astronomia. Normalment, les prediccions de les corbes de tendències es basen en observacions realitzades en el passat juntament amb un (sovint implícit) model del que l'observador creu que és el model subjacent del sistema observat. Els patrons de creixement exponencial són uns dels elements més freqüents en la natura, per la qual cosa cal tenir en compte en què solen convertir-se. Les quantitats regides per un creixement exponencial pur (la funció exponencial), Figura 1, sovint són molt elevades. Aquest tipus de creixement és una característica de molts sistemes joves. Fer prediccions a partir d'un model tan simple sovint porta a l'absurd. (Considereu les dimensions que tindria un humà de 18 anys si creixés al mateix ritme que durant el seu primer any de vida). Molt sovint, el que sembla ser un creixement exponencial s'estabilitzarà i s'acostarà a una asímptota horitzontal a mesura que vagi passant el temps (Figura 2). De fet, pel que sembla, la llei de Moore és en realitat una mena d'embolcall que cobreix les corbes de desenvolupament individual d'algunes tecnologies concretes (Figura 3). Com sabem que continuarem solucionant els problemes prou ràpidament per mantenir aquest embolcall exponencial? (El pessimista també recorda que la natura a vegades maleeix el creixement inicial amb un col·lapse catastròfic, com quan desapareix el subministrament alimentari i la població s'extingeix (Figura 4)). La majoria d'investigadors coincideixen a afirmar que no hi haurà cap obstacle tècnic insalvable que impedeixi que la llei de Moore segueixi essent vigent durant deu o quinze anys més. Més enllà d'aquest moment, és possible que es produeixi una ruptura pel que fa a la millora de les tècniques de fotolitografia que han impulsat la llei de Moore; però hi ha tecnologies radicals que poden impulsar la millora del hardware a partir d'aleshores. Així, molta gent creu que les millores del hardware a les quals fa referència la llei de Moore podrien continuar d'alguna manera generalitzada durant vint anys més, com a mínim. En aquest article, parteixo del supòsit que la llei de Moore pot ser vàlida durant vint o trenta anys més i plantejo quines poden ser les conseqüències d'aquest progrés. Com en el passat, podem sentir la temptació provocadora i dir, "tanta potència no és necessària". Però crec que, de fet, hi ha una aplicació magnífica que espera tota aquesta potència de hardware. Un dels motius que han inspirat aquesta conclusió és l'estimació de Hans Moravec de la potència de càlcul de la neurona [17], juntament amb la seva interpretació de la llei de Moore en relació amb aquesta estimació biològica [N2]. Aquest gràfic mostra que la paritat del hardware amb el cervell humà es produiria cap a l'any 2020. En certa manera, aquesta paritat no és el més espectacular [11]. L'especulació veritablement inquietant és imaginar què passarà després que s'hagi aconseguit la paritat: la llei de Moore segueix avançant. El hardware dels ordinadors aviat seria superior al cervell humà. En aquest article plantejo l'argument que aquest esdeveniment implicaria probablement la creació imminent, per part de la tecnologia, d'entitats amb una intel·ligència superior a la humana. L'esdeveniment més recent comparable és l'aparició de la vida humana a la Terra. El 1965, I. J. Good va escriure [11]:

"Definim una màquina ultraintel·ligent com una màquina capaç de superar de llarg totes les activitats intel·lectuals de qualsevol humà, per intel·ligent que sigui. Atès que el disseny de màquines és una d'aquestes activitats intel·lectuals, una màquina ultraintel·ligent podria dissenyar màquines fins i tot millors; sens dubte, hi hauria una "explosió d'intel·ligència" i la intel·ligència humana quedaria molt endarrerida... Per tant, la primera màquina ultraintel·ligent és la darrera invenció que l'home necessita, ja que la màquina és prou dòcil per dir-nos com mantenir-la sota control".

Good ha captat l'essència d'aquesta cursa cap endavant, però aquí tria no continuar amb les seves conseqüències més inquietants. Cap màquina intel·ligent del tipus que ell descriu seria una "eina" del humans, de la mateixa manera que els humans no són les eines dels conills, els pit-roigs o els ximpanzés. Quan és una intel·ligència superior a la humana qui impulsa el progrés, aquest progrés serà molt més ràpid. La millor analogia que trobo està en el passat evolutiu: els animals poden adaptar-se als problemes i fer invencions, però sovint no ho fan tan ràpidament com la selecció natural; el món actua com el seu propi simulador en el cas de la selecció natural. Els humans tenim la capacitat d'interioritzar el món i de desenvolupar escenaris hipotètics a les nostres ments; podem resoldre la majoria de problemes milers de vegades més ràpidament que la selecció natural. Ara bé, amb la creació dels mitjans per fer aquestes simulacions a unes velocitats molt més elevades, ens estem endinsant en un règim tan radicalment diferent del nostre passat humà, com el que va diferenciar els humans dels animals inferiors. Crec que podem afirmar que aquest esdeveniment és una singularitat (la "singularitat tecnològica" és el títol d'aquest article). És el moment en què els nostres antics models han de ser descartats i passa a governar una nova realitat. A mesura que avancem cap a aquest punt, influirà cada vegada més en els afers humans fins que la idea es converteixi en un tòpic. Als anys cinquanta hi havia molt pocs que ho afirmaven: Stan Ulam [28] parafrasejava les paraules de John von Neumann:

"Una conversa es centrava en el progrés cada vegada més ràpid de la tecnologia i en els canvis en la forma de vida humana, que suggereix l'apropament d'una singularitat essencial en la història de la raça humana més enllà de la qual la vida humana, tal i com la coneixem, no té continuïtat".

Von Neumann fins i tot utilitza el terme "singularitat", tot i que sembla que fa referència a un progrés normal i no a la creació d'una intel·ligència superior a la humana. (Per mi, la superhumanitat és l'essència de la singularitat. Sense això, tindríem una saturació de riqueses tècniques, mai adequadament assimilades (vegeu [25])).

Al llarg dels anys seixanta, setanta i vuitanta, es va difondre el reconeixement del cataclisme [11] [29] [1] [31] [5]. Potser van ser els escriptors de ciència-ficció els que van notar el primer impacte concret. Al cap i a la fi, els escriptors "durs" de ciència-ficció són els que proven d'escriure històries concretes sobre tot allò que la tecnologia pot fer per nosaltres. Encara més, aquests escriptors van anar veient gradualment un mur opac que travessa el futur. Hi va haver un temps en què podien situar les seves fantasies a milions d'anys en el futur [24]. Ara veuen que les seves extrapolacions més diligents abocaven a l'incognoscible en un termini molt curt. Hi va haver un temps en què els imperis galàctics podrien haver semblat un domini posthumà. Ara, per desgràcia, fins i tot ho són els interplanetaris.

Què passarà a les dues primers dècades del segle XXI, a mesura que anem apropant-nos cap al límit? Com es difondrà l'enfocament de la singularitat en la visió del món humà? Durant un temps encara, els detractors de la sagacitat de les màquines tindran bona premsa. Després de tot, fins que no tinguem un hardware tan potent com un cervell humà, probablement sigui una bogeria pensar que serem capaços de crear una intel·ligència equivalent a la humana (o fins i tot superior). (Hi ha la possibilitat que puguem aconseguir un equivalent humà a partir d'un hardware menys potent, sobretot si estem disposats a conformar-nos amb una ment artificial que sigui literalment lenta [30]. Però és molt més probable que dissenyar-ne el software sigui un procés difícil, que suposarà molts inicis i experimentacions falsos. Si és així, aleshores potser l'arribada de màquines conscients de si mateixes no es produeixi fins després del desenvolupament de hardware que sigui clarament superior a l'equipament natural dels humans).

Però, a mesura que passa el temps, hauríem de veure'n més símptomes. El dilema en què es troben els escriptors de ciència-ficció es percebrà en altres esforços creatius. (Alguns autors de còmics seriosos han expressat la seva preocupació sobre com aconseguir efectes espectaculars quan qualsevol cosa visible pot ser produïda mitjançant les tecnologies mes ordinàries). Veurem com l'automatització substituirà llocs de treball d'un nivell cada cop més elevat. Ja tenim eines (programes de matemàtiques simbòliques, cad/cam) que ens alliberen de la majoria de tasques de nivell inferior. O, dit d'una altra manera, la feina que és realment productiva està en mans d'una fracció de la humanitat cada vegada més petita i elitista. Amb l'arribada de la singularitat, veiem com les prediccions d'una desocupació autènticament tecnològica finalment es fan realitat.

Un altre símptoma de l'evolució cap a la singularitat: les idees s'haurien de difondre encara més de pressa i fins i tot les més radicals es convertirien ràpidament en tòpics. Quan vaig començar a escriure ciència-ficció a la dècada dels seixanta, semblava molt senzill trobar idees que tardarien dècades a penetrar en la consciència cultural; ara el termini necessari és d'uns divuit mesos. (És clar que això es podria deure només al fet que a mesura que em faig gran vaig perdent imaginació, però veig aquest efecte també en el altres.) Com una descàrrega en un flux compressible, la singularitat s'acosta a mesura que accelerem la velocitat crítica.

I què podem dir de l'arribada de la singularitat? Què es pot dir de la seva aparició? Ja que implica una cursa intel·lectual cap endavant, l'esdeveniment que la precipitaria podria ser inesperat; potser fins i tot per als investigadors implicats. ("Però tots els nostres models anteriors eren catatònics! Ens limitàvem a manipular alguns paràmetres...").

Es pot evitar la singularitat? (Quan hi ha un escenari clar i ben argumentat, sempre és divertit imaginar que aquesta escenari no es donarà mai i que ens encarreguen, potser d'aquí cinquanta anys, d'escriure un treball sobre per què era obvi que el futur previst no tenia sentit). Se m'acuden tres arguments a favor de la possibilitat d'un futur sense singularitat:

Potser els governs del món decidiran que la possibilitat és tan perillosa que la recerca que porta a la singularitat serà prohibida. Desgraciadament, el problema de la proliferació nuclear ja ha demostrat com n'és de fràgil aquesta esperança. Encara que tots els governs del món entenguessin l'"amenaça" i la temessin, l'evolució cap a l'objectiu continuaria. En la ficció, s'han escrit històries sobre l'aprovació de lleis que prohibien la construcció d'una "màquina semblant a la ment humana" [13]. De fet, l'avantatge competitiu -econòmic, militar, fins i tot artístic- de cada avenç en automatització és tan convincent que aprovar lleis que prohibeixin aquestes coses només assegura que algú altre les aconseguirà abans. Si pot haver-hi singularitat tecnològica, n'hi haurà. També és probable que els tecnooptimistes hagin senzillament subestimat el poder computacional del cervell humà o wetware. L'agost del 1992, Thinking Machines Corporation va organitzar un taller titulat "Com construirem una màquina que pensi" [27] per investigar la qüestió. Com podeu deduir pel títol del taller, els participants no donaven gaire suport als arguments en contra de la intel·ligència de les màquines. De fet, hi havia un acord general sobre el fet que les ments poden existir en substrats no biològics. Tanmateix, hi va haver un gran debat sobre la potència del hardware present en els cervells biològics. Una minoria creia que els més potents ordinadors de 1992 eren a tres ordres de magnitud de la potència del cervell humà. La majoria dels participants estaven d'acord amb l'estimació de Moravec [17] sobre el fet que estem a entre deu i quaranta anys de la igualtat del hardware. I encara hi havia una minoria va assenyalar [7][21] i va conjecturar que la competència computacional de les neurones simples podia ser molt més elevada del que es creia generalment. Si això fos cert, el hardware dels nostres ordinadors actuals podria ser d'uns deu ordres de magnitud inferior de l'equipament que duem instal·lat als nostres caps, i l'arribada de la singularitat per tant quedaria ajornada.

En la meva opinió, l'argument més probable contra la singularitat deriva d'una qüestió que la majoria de professionals de la ciència informàtica conceptuen com a "complexitat del software". Tot i que fem màquines que tinguin la potència del hardware humà, potser mai no entendrem com connectar-ne les parts correctament. Veritablement, la nostra capacitat per a crear grans programes eficaços no ha progressat, ni molt menys, amb la mateixa velocitat que la llei de Moore. Potser hi ha una cosa semblant al següent, una mica capriciós, contrapunt de Murphy a la llei de Moore:

La màxima eficàcia possible d'un sistema de software augmenta en proporció directa al logaritme de l'eficàcia (és a dir, velocitat, amplada de banda, capacitat de memòria) del hardware subjacent.

Si la complexitat del hardware triomfés, aleshores a principis del segle vint-i-un trobaríem que les corbes de rendiment del nostre hardware comencen a estabilitzar-se, això a causa de la nostra incapacitat per subministrar el suport programat per a la recerca de millores posteriors del hardware. Acabaríem amb un hardware potentíssim, però sense la capacitat de fer-lo anar més lluny. Res no "es despertaria" mai; mai no hi hauria la cursa intel·lectual cap endavant que és l'essència de la singularitat. Podria haver-hi una edat d'or a l'estil de Gunther Stent. (De fet a la pàgina 137 de [25], Stent cita explícitament el desenvolupament de la intel·ligència superhumana com una condició suficient per trencar les seves projeccions). Però també significaria la fi del progrés. El progrés pel que fa a la memòria dels ordinadors probablement seria l'últim a estabilitzar-se, ja que el xips de RAM són unes estructures molt regulars, i els ordinadors portàtils del futur poden fàcilment contenir tot el software escrit pels humans. En aquest futur, l'arqueologia del software seria un substitut comú per a la programació nova. Des del meu punt de vista, la possibilitat d'un llegat de software amb mil·lenis d'antiguitat no és gaire atractiva [N9].

Si la singularitat tecnològica és el nostre futur, com serà? Bé, jo l'anomeno singularitat en part perquè és una pregunta difícil de contestar. Pot ser divertit especular, però és una mica com si un cap-gros especulés sobre el futur de l'administració dels aiguamolls. Analitzem la part més fosca: fins a quin punt podria ser nefasta l'era posthumana? Bé... força nefasta. L'extinció física de la raça humana n'és una possibilitat. (O com Eric Drexler va dir sobre la nanotecnologia: si es tingués en compte tot el que aquesta tecnologia pot fer, potser els governs senzillament decidirien que ja no necessiten mai més ciutadans!). Tanmateix, potser l'extinció física és la possibilitat més temuda. Un altre cop, analogies: penseu en les diferents maneres que tenim de relacionar-nos amb els animals. Alguns dels abusos físics més crus són inversemblants, i tot i així... En un món posthumà encara hi haurà molts camps on l'automatització de l'equivalent humà seria desitjable: sistemes incorporats a aparells autònoms, daemons conscients de si mateixos que actuarien en el fons de consciències més grans. (Una intel·ligència superhumana podria ser una Societat de la Ment [16] amb alguns components molt competents). Alguns d'aquests equivalents humans podrien ser usats només per al processament de senyals digitals. Podrien assemblar-se més a balenes que a humans. Altres podrien assemblar-se molt als humans, però encara amb una dedicació exclusiva que en els nostres temps els faria ingressar en un hospital psiquiàtric. Tot i que cap d'aquestes criatures podrien ser humans de carn i ossos, sí que podrien ser la cosa que més s'hi assembla en el nou medi que ara anomenem humà. (I. J. Good va dir alguna cosa sobre això, tot i que avui en dia el seu consell pot ser discutible: Good [12] va proposar una "metanorma d'or", que podria parafrasejar-se com "tracta els teus inferiors com t'agradaria que els teus superiors et "tractessin". És una idea meravellosa i paradoxal (i molts dels meus amics no se la creuen) ja que els beneficis de la teoria dels jocs són molt difícils d'articular. Tanmateix si fóssim capaços de seguir la regla de Good, en cert sentit això podria dir alguna cosa sobre la versemblança d'una amabilitat com aquesta en aquest univers). He argumentat que si la singularitat és possible, aleshores no podem evitar-la; la seva arribada serà una conseqüència inevitable de la competitivitat natural dels humans i de les possibilitats inherents en la tecnologia. I a més... nosaltres en som els iniciadors. Si la transició dura dècades, el que jo anomeno una "arrencada suau", podem tenir temps de planejar-la i orientar de manera adequada els nous poders. Per desgràcia, la transició pot ser molt ràpida, potser només un centenar d'hores, una "arrencada difícil". (A [5], Greg Bear fa un retrat dels canvis més importants que succeeixen en unes quantes hores.) Seria molt complicat planificar una arrencada difícil. Si tenim en compte aquest progrés, podria ajudar-nos considerar diverses vies de recerca diferents (fins i tot si la realitat potser depèn d'una certa barreja d'enfocaments):

Quan la gent parla de crear éssers amb una intel·ligència superhumana, sovint s'imagina un projecte d'intel·ligència artificial. Però hi ha altres vies. La recerca en xarxes informàtiques i interfícies entre humans i màquines podrien conduir a la singularitat. Anomeno aquest enfocament oposat amplificació de la intel·ligència. L'amplificació de la intel·ligència és alguna cosa que avança de manera molt natural; en la majoria de casos no és ni tan sols reconeguda

per aquells que la desenvolupen com el que és. Però cada vegada que millorem la nostra capacitat per accedir a la informació i per comunicar-la als altres, en certa manera millorem la intel·ligència natural. Fins i tot ara, l'equip format per un humà amb un doctorat i una bona estació de treball (fins i tot sense connexió a la xarxa!) probablement podria treure la màxima puntuació en qualsevol test d'intel·ligència que se li pogués fer.

I passa una cosa molt semblant amb el fet que l'amplificació de la intel·ligència és un camí molt més senzill per aconseguir la superhumanitat que la intel·ligència artificial pura. En els humans, els problemes més difícils de desenvolupar ja s'han solucionat. Fer-se més forts a partir d'uns mateixos hauria de ser més fàcil que entendre primer què som en realitat i després construir màquines que siguin tot això. I hi ha com a mínim un precedent basat en conjectures per a aquest enfocament. Cairns-Smith [6] ha especulat sobre el fet que la vida biològica pot haver començat com un apèndix d'una vida encara més primitiva basada en el creixement cristal·lí. Lynn Margulis (a [15] i a d'altres llocs) ha elaborat uns arguments consistents per defensar que el mutualisme és una enorme força impulsora de l'evolució. Fixeu-vos que no estic proposant que la investigació en el terreny de la intel·ligència artificial s'hagi d'ignorar o deixar de subvencionar. El que s'esdevingui en el camp de la intel·ligència artificial sovint tindrà aplicacions en amplificació de la intel·ligència, i viceversa. Proposo que reconeguem que en la recerca de xarxes i de interfícies hi ha alguna cosa tan profunda (i en potència desbordant) com la intel·ligència artificial. Des d'aquesta perspectiva, podem veure projectes que no són tan directament aplicables com la interfície convencional i el treball de disseny de xarxes, però que serveixen per fer-nos avançar cap a la singularitat en el camí de l'amplificació de la intel·ligència.

Aquests són alguns dels projectes possibles que prenen una importància especial, segons el punt de vista de l'amplificació de la intel·ligència:

· Desenvolupament de la simbiosi humà/ordinador en l'art i combinació de la capacitat de generació de gràfics de les màquines modernes i la sensibilitat estètica dels humans. És clar que s'ha investigat molt en el disseny d'assistents informàtics per als artistes, com a eines que estalvien feina. Proposo que explícitament aspirem a una fusió més gran de la competència, que explícitament reconeguem l'enfocament cooperatiu que és possible. Karl Sims [23] ha treballat molt en aquesta direcció.

· Permetre els equips d'humans/ordinadors als torneigs d'escacs. Ja tenim programes que poden jugar millor que la majoria dels humans. Però quant s'ha treballat sobre com podria utilitzar un humà aquesta potència per aconseguir alguna cosa fins i tot millor? Si es permetessin aquests equips en, com a mínim, alguns torneigs d'escacs, es podria aconseguir un efecte positiu en la recerca en amplificació de la intel·ligència com el que va tenir l'entrada d'ordinadors en els torneigs per a la intel·ligència artificial.

· Explotar Internet com una eina que combina l'humà amb la màquina. De tots els punts de la llista, aquest és el que està avançant més ràpidament i que pot fer topar amb la singularitat abans que cap altre. El poder i la influència de fins i tot l'Internet actual han estat molt subestimats. Per exemple, crec que els nostres sistemes informàtics contemporanis es trencarien sota el pes de la seva pròpia complexitat si no fos per l'avantatge que la "ment de grup" de la xarxa dóna a l'administració del sistema i a la gent que li dóna suport! L'autèntica anarquia de la xarxa mundial és una evidència del seu potencial. Mentre augmenta la connectivitat, l'amplada de banda, la mida dels arxius i la velocitat de l'ordinador, ens trobem davant d'una cosa semblant a la visió que tenia Lynn Margulis [15] de la biosfera com un processador de dades resumit, però un milió de vegades més ràpid i amb milions d'agents humanament intel·ligents (nosaltres mateixos). Curiosament, el resultat podria ser superhumà sense ser conscient de sí mateix [N5][N6]. (Una versió extrema d'aquesta trajectòria d'Internet cap a la singularitat podria traduir-se en una connexió de xarxa molt fragmentada [N10], que potser conduiria a una situació on el medi en sí mateix té una certa intel·ligència [N3]).

Els exemples anteriors il·lustren la recerca que es pot fer en el context de la ciència informàtica contemporània. Hi ha altres paradigmes. Per exemple, molt del treball que es fa en intel·ligència artificial i en xarxes neuronals es beneficiarien d'una relació més estreta amb la vida biològica. En lloc de simplement provar d'imitar i entendre la vida biològica mitjançant els

ordinadors, les investigacions podrien dirigir-se cap a la creació de sistemes compostos que es basin en la vida biològica per orientar o proporcionar característiques que no entenem prou bé encara per implementar en el hardware. Un dels somnis més antics de la ciència-ficció ha estat la connexió directa d'un cervell a interfícies d'ordinadors [2] [29]. De fet, hi ha unes tasques concretes que es poden fer (i que s'estan fent) en aquesta àrea:

· Les pròtesis tenen una aplicació comercial directa. Es poden fer nervis cap a sensors de silicona [14]. Aquest és un pas fascinant a curt termini cap a la comunicació directa.

· Les connexions directes als cervells semblen factibles, si la velocitat de transmissió de dades és lenta: si tenim en compte la flexibilitat de l'aprenentatge humà, potser no caldria seleccionar amb exactitud els objectius de les neurones cerebrals. Fins i tot 100 bits per segon serien una gran ajuda per a les víctimes d'embòlia cerebral, que d'altra manera es veurien limitades a utilitzar interfícies de menú.

· La connexió a la línia òptica té un gran potencial per aconseguir amplades de banda d'1 Mbit per segon, més o menys. Però per fer-ho necessitem conèixer l'arquitectura de visió a una escala molt fina, i necessitem situar una enorme xarxa d'elèctrodes amb una gran precisió. Si volem que la nostra connexió a un gran ample de banda molt s'afegeixi a les trajectòries que ja hi ha al cervell, el problema es fa molt més difícil. No serà suficient que implantem una malla de receptors de gran ample de banda dins un cervell. Però imagineu que la malla de gran ample de banda ja hi fos mentre l'estructura del cervell s'anava configurant, mentre l'embrió es desenvolupa. Això suggereix:

· Experiments amb embrions d'animals. Jo no esperaria cap èxit de l'amplificació de la intel·ligència els primers anys d'aquest tipus de recerca, però per a les persones que estudien com es desenvolupa el cervell en estat embrionari podria ser molt interessant ja que el cervells que s'estan desenvolupant donen accés a complexes estructures neuronals simulades. A llarg termini, aquests experiments podrien produir animals amb trajectòries sensorials addicionals i capacitats intel·lectuals interessants.

Inicialment, tenia esperances que aquest comentari sobre l'amplificació de la intel·ligència aportés alguns plantejaments clarament més segurs pel que fa a la singularitat. (Al cap i a la fi, l'amplificació de la intel·ligència permet la nostra participació en una mena de transcendència). Malauradament, en tornar a revisar aquestes propostes d'amplificació de la intel·ligència, de tot el que estic segur és que haurien de tenir-se en compte i que ens poden donar més opcions. Però pel que fa a la seguretat... bé, algunes de les coses suggerides poden intimidar una mica. Un dels meus crítics va assenyalar que l'amplificació de la intel·ligència en individus humans genera una elit força sinistra. Els humans tenim milions d'anys de bagatge evolutiu que ens fa contemplar la competència d'una manera molt crítica. Gran part d'aquesta actitud potser no és necessària en el món actual, en què els perdedors adopten els trucs dels guanyadors i s'adhereixen a les seves empreses. Una criatura que hagués estat construïda de nou probablement podria ser una entitat molt més benigna que una amb un origen basat en els ullals i les urpes. I fins i tot la visió igualitària d'una Internet que es desperta amb tota la humanitat pot considerar-se com un malson [26].

El problema no és simplement que la singularitat representa la desaparició de la humanitat del centre de l'escenari, sinó que contradiu les nostres nocions més profundament arrelades sobre l'ésser. Crec que una mirada atenta a la perspectiva més optimista pot mostrar per què és així: suposem que aconseguim una arrencada suau i que som capaços d'adaptar la singularitat a mida. Suposem que podem ser-ne participants, i que nosaltres mateixos controlem l'exponencial. Suposem que podem fer realitat les nostres esperances més extravagants. Aleshores el que demanaríem seria que els humans mateixos es convertissin en els seus propis successors, que qualsevol tipus d'injustícia que es produís fos atenuada pel coneixement que tenim de les nostres arrels. Per a aquells que no es veiessin afectats, l'objectiu seria un tractament benigne (potser fins i tot fer que els enrederits semblessin els amos d'uns esclaus divins). Podria ser una edat d'or que també impliqués progrés (saltant la barrera de Stent). La immortalitat seria possible (o, si més no, una vida que durés tant com l'univers[10] [4]).

Però en aquest món tan prometedor i amable, els problemes filosòfics són intimidatoris. Una ment que es manté en la mateixa capacitat no pot viure per sempre; després d'uns quants milers d'anys semblaria més una gravació repetitiva que una persona. (El quadre més esgarrifós que he vist sobre això és a [18].) Per viure indefinidament, la ment hauria de créixer... I quan fos suficientment gran, i mirés enrere... quina afinitat podria tenir amb l'ànima que era originàriament? Segurament l'ésser posterior seria tot el que era l'original, però molt més. I passaria el mateix fins i tot en el cas de l'individu. Aleshores seria vàlida la noció de Cairns-Smith d'una vida nova que creix de forma incremental a partir de la vella.

Aquest "problema" de la immortalitat es presenta de maneres molt més directes. La noció d'ego i de consciència d'un mateix ha estat la base del racionalisme pràctic dels darrers segles. Encara ara la noció consciència d'un mateix és atacada pels entusiastes de la intel·ligència artificial ("consciència d'un mateix i altres falses il·lusions"). L'amplificació de la intel·ligència debilita el nostre concepte d'ego des d'una altra direcció. El món posterior a la singularitat suposarà una gran activitat de xarxes de gran ample de banda. Una característica important de les entitats molt superiors a les humanes serà segurament la seva capacitat per comunicar-se en amples de banda variables, incloses alguns molt superiors a la parla o als missatges escrits. Què passa quan les peces del jo es poden copiar i fusionar, quan la mida d'una consciència d'un mateix pot créixer o encongir-se per adaptar-se a la naturalesa dels problemes que s'analitzen? Sospito que aquestes siguin característiques essencials de la singularitat [N8]. Si pensem en aquests temes, començarem a intuir com serà d'essencialment estranya i diferent l'era posthumana per molt hàbil i benigna que sigui la forma en què es duu a terme el procés de canvi.

D'una banda, la visió encaixa amb molts dels nostres somnis més feliços: un temps sense fi, on realment ens puguem conèixer i on puguem entendre els misteris més insondables. D'altra banda, s'assembla molt al pitjor dels escenaris que he imaginat en començar aquesta ponència.

Quin és el punt de vista vàlid? De fet, crec que la nova era simplement és massa diferent perquè encaixi en el nostre marc clàssic. Aquest marc es basa en la idea de ments aïllades i immutables connectades mitjançant dèbils enllaços amb poc ample de banda. Però el món posterior a la singularitat sí encaixa amb la llarga tradició de canvi i cooperació que va començar fa molt temps (potser fins i tot abans que aparegués la vida biològica). Crec que hi ha nocions d'ètica que es podrien aplicar en una era com aquesta. La recerca de l'amplificació de la intel·ligència i les comunicacions amb gran ample de banda alta haurien de millorar aquesta comprensió. Ara només podem albirar-ho. Hi ha la "metanorma d'or" de Good; potser hi ha regles per diferenciar un mateix dels altres sobre la base de l'ample de banda ampla de la connexió. I tot i que la ment i el jo seran molt més inestables del que ho van ser en el passat, no caldrà que es perdin moltes de les coses que valorem (coneixement, memòria, pensament). Penso que Freeman Dyson té raó quan diu [9]: "Déu és en el que es converteix la ment quan va més enllà de l'escala del nostre enteniment".

==========

[Al llarg dels anys, molta gent -més de la que puc esmentar- ha col·laborat en el meu debat gràcies als seus comentaris.
El treball original es va beneficiar de les discussions amb John Carroll de la San Diego State University i amb Howard Davidson de Sun Microsystems].

Fonts esmentades [i sol·licitud ocasional d'ajut bibliogràfic]

[1] Alfvén, Hannes, sota el pseudònim d'Olof Johanneson, The End of Man?, Award Books, 1969 anteriorment publicat com "The Tale of the Big Computer", Coward-McCann, traduït d'un llibre amb copyright de 1966 d'Albert Bonniers Forlag AB, amb copyright sobre la traducció anglesa del 1966 de Victor Gollanz, Ltd.

[2] Anderson, Poul, "Kings Who Die", If, març 1962, pàgs. 8-36. Reeditat a Seven Conquests, Poul Anderson, MacMillan Co., 1969.

[3] Asimov, Isaac, "Runaround", Astounding Science Fiction, març 1942, pàg. 94. Reeditat a Robot Visions, Isaac Asimov, ROC, 1990. Les lleis de la robòtica d'Asimov. La versió resumida d'aquesta idea és la següent: cal construir robots amb instints que impedeixen que facin mal als humans. Aquestes lleis poden dur-nos a la singularitat sense grans riscos. Personalment, crec que qualsevol regla prou estricta per ser eficaç també produiria un artefacte les capacitats del qual serien clarament inferiors a les de versions sense restriccions (i,per tant, la competència humana afavoriria el desenvolupament dels models més perillosos). Tot i així, el somni d'Asimov és meravellós: imagineu-vos un esclau servicial, que té 1.000 vegades més aptituds que vosaltres en tots els sentits. Imagineu una criatura que pogués satisfer tots i cadascun dels vostre desigs segurs (sigui el que sigui el que això signifiqui) i que encara tingués el 99,9% del seu temps lliure per a altres activitats. Hi hauria un univers nou que nosaltres mai no entendríem realment, però ple de déus benèvols (tot i que un dels meus desigs personals seria convertir-me en un d'ells).

[4] Barrow, John D. i Frank J. Tipler, The Anthropic Cosmological Principle, Oxford University Press, 1986.

[5] Bear, Greg, "Blood Music", Analog Science Fiction-Science Fact, juny, 1983. Ampliat a la novel·la Blood Music, Morrow, 1985.

[6] Cairns-Smith, A. G., Seven Clues to the Origin of Life, Cambridge University Press, 1985.

[7] Conrad, Michael et al., "Towards an Artificial Brain", BioSystems, vol. 23, pàgs. 175-218, 1989.

[8] Drexler, K. Eric, Engines of Creation, Anchor Press/Doubleday, 1986.

[9] Dyson, Freeman, Infinite in All Directions, Harper & Row, 1988.

[10] Dyson, Freeman, "Physics and Biology in an Open Universe", Review of Modern Physics, vol. 51, pàg. 447-460, 1979.

[11] Good, I. J., "Speculations Concerning the First Ultraintelligent Machine", a Advances in Computers, vol. 6, Franz L. Alt i Morris Rubinoff, editors, pàgs. 31-88, 1965, Academic Press.

[12] Good, I. J., [Ajudeu-me! No puc trobar la font de la "metaregla d'or" de Good, tot i que recordo clarament haver-ne sentit a parlar cap als anys seixanta. Gràcies a l'ajuda de la xarxa, he trobat referències d'una gran quantitat de temes relacionats. G. Harry Stine i Andrew Haley han escrit sobre la metallei ja que podrien aplicar-se als extraterrestres: G. Harry Stine, "How to Get along with Extraterrestrials ... or Your Neighbor", Analog Science Fact-Science Fiction, febrer, 1980, pàgs. 39-47.]

[13] Herbert, Frank, Dune, Berkley Books, 1985. Tanmateix, aquesta novel·la es va publicar en capítols a Analog Science Fiction-Science Fact als anys seixanta.

[14] Kovacs, G.T.A. et al., "Regeneration Microelectrode Array for Peripheral Nerve Recording and Stimulation", IEEE Transactions on Biomedical Engineering, vol. 39, núm. 9, pàgs. 893-902.

[15] Margulis, Lynn i Dorion Sagan, Microcosmos, Four Billion Years of Evolution from Our Microbial Ancestors, Summit Books, 1986.

[16] Minsky, Marvin, Society of Mind, Simon and Schuster, 1985.

[17] Moravec, Hans, Mind Children, Harvard University Press, 1988.

[18] Niven, Larry, "The Ethics of Madness", If, abril 1967, pàgs. 82-108. Reeditat a Neutron Star, Larry Niven, Ballantine Books, 1968.

[19] Penrose, Roger, The Emperor's New Mind, Oxford University Press, 1989. Un argument contundent contra la possibilitat que les màquines tinguin consciència.

[20] Platt, Charles, Comunicació privada. Quan estava escrivint el treball al 1993, Charles Platt va assenyalar que durant 30 anys alguns entusiastes de l'amplificació de la intel·ligència havien fet afirmacions extravagants sobre "els propers trenta anys". Per tal de no ser culpable d'una ambigüitat temporal d'aquest tipus, vaig escriure que em sorprendria que la singularitat es produís abans del 2005 o després del 2030.

[21] Rasmussen, S. et al., "Computational Connectionism within Neurons: a Model of Cytoskeletal Automata Subserving Neural Networks", a Emergent Computation, Stephanie Forrest, ed., pàgs. 428-449, MIT Press, 1991.

[22] Searle, John R., "Minds, Brains, and Programs", a The Behavioral and Brain Sciences, vol. 3, Cambridge University Press, 1980. Un argument contundent contra la possibilitat que les màquines puguin arribar a tenir consciència. El treball de Searle es va reeditar a The Mind's I, editat per Douglas R. Hofstadter i Daniel C. Dennett, Basic Books, 1981 (la meva font per a aquesta referència). Aquesta reedició conté una excel·lent crítica de l'article.

[23] Sims, Karl, "Interactive Evolution of Dynamical Systems", Thinking Machines Corporation, Technical Report Series (publicat a Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life, París, MIT Press, desembre de 1991).

[24] Stapledon, Olaf, The Starmaker, c. 1937 (hi ha una reedició recent de Penguin, 1988).

[25] Stent, Gunther S., The Coming of the Golden Age: A View of the End of Progress, The Natural History Press, 1969.

[26] Swanwick, Michael, Vacuum Flowers, publicat en capítols a Isaac Asimov's Science Fiction Magazine, desembre(?) 1986 - febrer 1987. Reeditat per Ace Books, 1988.

[27] Thearling, Kurt, "How We Will Build a Machine that Thinks", un taller a Thinking Machines Corporation, 24 a 26 d'agost de 1992. Comunicació personal.

[28] Ulam, S., Tribute to John von Neumann, Bulletin of the American Mathematical Society, vol. 64, núm. 3, part 2, maig de 1958, pàgs. 1-49.

[29] Vinge, Vernor, "Bookworm, Run!", Analog, març de 1966, pàgs. 8-40. Reeditat a The Collected Stories of Vernor Vinge, Tor Books, 2001.

[30] Vinge, Vernor, "True Names", Binary Star Number 5, Dell, 1981. Reeditat a True Names and the Opening of the Cyberspace Frontier, J. Frenkel, editor, Tor Books, 2001.

[31] Vinge, Vernor, "First Word", Omni, gener de 1983, pàg. 10.

Algunes referències addicionals afegides el 2002

[N1] Moore, Gordon E., "Cramming more components onto integrated circuits", Electronics, vol. 38, núm. 8, 19 d'abril de 1965, com s'assenyala a http://www.intel.com/research/silicon/mooreslaw.htm.
L'article d'Electronics té una il·lustració profètica d'un home que ven "ordinadors domèstics", en uns grans magatzems!

[N2] Moravec, Hans, Robot: Mere Machine to Transcendent Mind, Oxford University Press, 1999. Continuacuó de [17]. Vegeu també: http://www.transhumanist.com/volume1/moravec.htm

[N3] Schroeder, Karl, Ventus, Tor Books, 2001.

[N4] Sterling, Messina i Smith, Enabling Technologies for Petaflops Computing, MIT Press, 1995. Una mirada relativament no especulativa sobre la potència dels ordinadors que podrà aconseguir-se el 2014.

[N5] Stiegler, Marc, Earthweb, Baen Books, 1999.

[N6] Stock, Gregory, Metaman, Simon&Shuster, 1993.

[N7] Vinge, Vernor, "The Coming Technological Singularity: How to Survive in the Post-Human Era", 1993, disponible en format electrònic a http://www-rohan.sdsu.edu/faculty/vinge/misc/singularity.html.

[N8] Vinge, Vernor, "Nature, Bloody in Tooth and Claw?", 1996, disponible en format electrònic a http://www-rohan.sdsu.edu/faculty/vinge/misc/evolution.html

[N9] Vinge, Vernor, A Deepness in the Sky, Tor Books, 1999.

[N10] Vinge, Vernor, "Fast Times at Fairmont High", a The Collected Stories of Vernor Vinge, Tor Books, 2001.