

# La Singularité

Vernor Vinge

<http://www.mindstalk.net/vinge>

Département des Sciences Mathématiques

San Diego State University

Traduit de l'anglais par Emmanuel Mie

<http://www.alchimie-web.com>

1993 Vernor Vinge

Cet article peut être reproduit à des fins non commerciales  
à condition d'être copié dans son intégralité, y-compris cette note

La version originale de cet article a été présentée au

## Symposium VISION-21

parrainé par

le Lewis Research Center de la NASA

et

l'Ohio Aerospace Institute,

les 30 et 31 mars 1993.

Une version légèrement remaniée a paru dans  
le numéro de l'hiver 1993 de la revue *Whole Earth Review*.

## Résumé

D'ici trente ans, nous aurons les moyens de créer une intelligence superhumaine. Peu de temps après cela, l'ère humaine s'achèvera. Peut-on éviter une telle éventualité ? Si on ne peut l'éviter, peut-on orienter les événements de façon à assurer la survie de l'Humanité ? Ces questions font l'objet d'investigations. Voici quelques-unes des réponses possibles (et quelques risques connexes).

## Table des matières

<b>Qu'est-ce que la Singularité ?.....</b>	<b>3</b>
<b>La Singularité peut-elle être évitée ?.....</b>	<b>8</b>
<b>D'autres chemins vers la Singularité :</b>	
<b>l'Intelligence Amplifiée.....</b>	<b>11</b>
<b>La superhumanité forte et ce que nous pouvons attendre de mieux.....</b>	<b>17</b>
<b>Bibliographie annotée (et, au passage, un appel à candidats pour le travail bibliographique.....)</b>	<b>19</b>

## Qu'est-ce que la Singularité<sup>1</sup> ?

L'accélération du progrès technologique est une des caractéristiques centrales de ce siècle. Dans cet article, j'avance que nous sommes à la veille d'un changement comparable à l'émergence de la vie humaine sur Terre. La cause précise de ce changement est la création imminente d'entités technologiques possédant une intelligence supérieure à celle de l'humain. Il existe plusieurs chemins par lesquels la science pourrait réussir cette percée (et c'est une raison de plus pour être certain que cet événement va se produire) :

- » Ce pourraient être des ordinateurs évolués qui parviendraient à *l'éveil* et à l'intelligence superhumaine. (À ce jour, la question de savoir s'il est possible de créer une machine d'intelligence équivalente à l'humain reste controversée. Mais si la réponse est *oui*, il ne fait pas de doute que des êtres plus intelligents peuvent être construits peu de temps après.)
- » De vastes réseaux informatiques (et leurs utilisateurs associés) pourraient *s'éveiller* en tant qu'entités dotées d'intelligence superhumaine.
- » Les interfaces humain / ordinateur pourraient devenir si intimes qu'il ne serait pas déraisonnable d'en considérer les utilisateurs comme superhumainement intelligents.
- » La science biologique pourrait fournir des moyens d'améliorer les capacités intellectuelles naturelles de l'humain.

Les trois premières possibilités dépendent en grande partie du perfectionnement du matériel informatique. Le progrès en la matière a suivi une courbe remarquablement régulière ces dernières décades [17]. En m'appuyant sur cette tendance, je crois que l'émergence d'une intelligence supérieure à celle de l'humain aura lieu dans les trente

---

<sup>1</sup> <http://www.aleph.se/Trans/Global/Singularity/>

prochaines années. (Charles Platt [20] fait remarquer que les enthousiastes de l'IA\* ont émis de telles prévisions au cours des trente dernières années. Pour qu'on ne m'accuse pas d'ambiguïté temporelle, je vais être plus spécifique : je serais surpris si cet événement survenait avant 2005 ou après 2030.)

Quelles sont les conséquences d'un tel événement ? À partir du moment où une intelligence supérieure à celle de l'humain prendra les commandes du progrès, ce dernier va connaître une forte accélération. En fait, il ne semble pas y avoir de raison qui empêcherait le progrès de susciter la création d'entités encore plus intelligentes – et dans un temps toujours plus court. La meilleure comparaison que je puisse faire est celle de l'évolution naturelle dans le passé : les animaux sont capables de s'adapter aux problèmes et d'inventer des solutions, mais au rythme de la sélection naturelle ; dans le cas de la sélection naturelle, le monde constitue son propre simulateur. Nous autres humains, en revanche, sommes capables d'intérioriser le monde, de nous poser la question : « *Qu'est-ce qui se passe si... ?* » Cela nous permet de résoudre les problèmes des milliers de fois plus vite que par sélection naturelle. Ainsi, en créant les moyens d'exécuter ces simulations à des vitesses bien supérieures à celle de l'esprit humain, nous entrons dans un régime aussi éloigné de notre réalité humaine que les animaux inférieurs le sont de la nôtre.

Du point de vue humain, ce changement entraînera la remise en cause de toutes les règles existantes, presque instantanée, une fuite en avant à une vitesse exponentielle, hors de tout contrôle. Des développements dont on pensait qu'ils n'arriveraient pas avant *1 million d'années* (voire jamais) sont susceptibles d'arriver dans le prochain siècle. (Greg Bear dépeint le tableau des principaux changements qui adviendraient en quelques heures. [5])

Il me semble pertinent d'appeler cet événement une singularité\* (LA Singularité, pour notre propos). C'est le point où les vieux schémas de pensée doivent être écartés parce qu'une

---

\* Intelligence Artificielle (NdT)

\* En mathématiques, une *singularité* est en général un point, une valeur ou un cas dans lequel un certain objet mathématique n'est pas bien défini. Pour approfondir, [Wikipedia](#), évidemment ! (NdT)

nouvelle réalité prévaut. Au fur et à mesure que nous nous rapprochons de ce point, cette notion va s'immiscer de plus en plus intimement dans nos vies quotidiennes, jusqu'à devenir un lieu commun. Pourtant, quand nous l'atteindrons, nous risquons néanmoins d'être confrontés à une grande surprise, et à un inconnu plus grand encore. Dans les années 1950, ils étaient très peu nombreux à être conscients de tout cela ; Stan Ulam évoque une conversation avec John Von Neumann :

*[Cette] conversation [était] centrée sur l'accélération infinie des progrès de la technologie et leur influence sur les modes de vie humains, qui donnerait l'impression d'approcher une singularité essentielle dans l'histoire de la race au-delà de laquelle les considérations humaines actuelles n'auraient plus lieu d'être. [28]*

Von Neumann lui-même emploie le terme "singularité", bien qu'il semble parler du progrès normal et non de l'émergence d'un intellect superhumain. (Pour moi, le caractère superhumain est l'essence même de la Singularité. Sans ce caractère, nous ne parviendrons qu'à une surenchère d'avancées technologiques que nous resterons incapables d'assimiler. [25])

Dans les années 1960, quelques-unes des implications de l'intelligence superhumaine ont commencé à être perçues. L.J. Good écrit :

*Soit une machine ultraintelligente, à savoir une machine capable de surpasser de loin toutes les activités intellectuelles de n'importe quel humain, aussi intelligent soit-il. Puisque la conception de telles machines est une de ces activités intellectuelles, une machine ultraintelligente serait capable de concevoir des machines encore plus performantes ; on aboutirait alors indubitablement à une "explosion d'intelligence", qui laisserait lon derrière elle l'intelligence humaine. C'est pourquoi, la conception d'une telle machine est la dernière chose dont l'homme aura jamais besoin, à supposer que la machine soit suffisamment docile pour nous expliquer comment la garder sous notre*

*contrôle. Le plus probable est que, dans le courant du XX<sup>e</sup> siècle, une machine ultraintelligente sera construite, et que ce sera la dernière chose dont l'Humanité avait besoin.* [11]

Good a perçu l'essence de la fuite, mais les conséquences les plus troublantes lui échappent. Une machine telle qu'il la décrit ne saurait être un "outil" pour l'Humanité – pas plus que les humains ne sont les outils des lapins, des rouge-gorges ou des chimpanzés.

Entre les années 1960 et 1990, l'intuition du cataclysme est devenue plus courante [29] [1] [31] [5]. Ce sont sans doute les auteurs de Science-Fiction qui en ont ressenti les premiers symptômes concrets. Après tout, les auteurs de Science-Fiction *pure et dure* s'efforcent de décrire tout ce que la technologie pourrait faire pour nous. De plus en plus, ces auteurs se sont heurtés à un mur opaque qui barre le futur. Auparavant, ils pouvaient situer leurs fictions à des millions d'années dans le futur [24]. Dorénavant, ils s'aperçoivent que leurs extrapolations les plus poussées résultent dans l'inconnu... à court terme. Ils pouvaient imaginer la post-Humanité loin dans le futur, au temps des empires galactiques. Dans la Science-Fiction d'aujourd'hui, tristement, les post-Humains\* sont cantonnés aux empires interplanétaires.

Qu'en est-il des années 1990 à 2010, qui nous rapprochent de la limite ? Comment l'Humanité va-t-elle prendre conscience de l'imminence de la Singularité ? Pendant un temps, les critiques habituelles sur la sagesse des machines vont continuer de fleurir. Après tout, tant que nous ne disposons pas de matériel aussi puissant que le cerveau humain, il peut sembler idiot d'imaginer que nous serons un jour capables de produire une intelligence équivalente (ou supérieure) à celle de l'homme. (Il y a bien la possibilité un peu tirée par les cheveux de créer un équivalent avec moins de puissance matérielle, en acceptant de laisser la vitesse de côté, en acceptant de créer un être artificiel littéralement lent [30]. Mais il est bien plus probable que le codage du logiciel sera un processus délicat,

---

\* La post-Humanité est un concept issu de la Science-Fiction et de la philosophie transhumaniste. Un post-Humain est un être hypothétique du futur, "*dont les aptitudes de base dépassent si radicalement celles des humains actuels qu'on ne peut plus le considérer comme humain d'après nos standards actuels.*" ([The Transhumanist FAQ](#))  
Pour approfondir : [Wikipedia](#) (l'article en anglais me semble plus pertinent) (NdT)

plein de faux départs et de tâtonnements. S'il en est ainsi, des machines conscientes d'elles-mêmes ne pourront émerger qu'après le développement de matériel substantiellement plus puissant que l'équipement naturel de l'humain.)

Pourtant, avec le temps, nous devrions voir apparaître de plus en plus de symptômes. Le dilemme ressenti par les auteurs de Science-Fiction va percer dans d'autres domaines créatifs. (J'ai entendu des auteurs de BD perplexes se poser la question : comment créer des effets spectaculaires quand des outils technologiques à portée de tous sont capables de produire tout ce qu'il est possible de voir ?) L'automate va remplacer l'humain à des niveaux de qualification toujours plus élevés. Nous disposons déjà d'outils (programmes de mathématiques symboliques, DAO, MAO) qui nous libèrent d'une grande partie des tâches les moins gratifiantes. Autrement dit, le travail réellement productif est le domaine d'une proportion de plus en plus réduite et de plus en plus qualifiée de l'Humanité. La proximité de la Singularité rend crédibles les prédictions d'un véritable chômage technologique.

Un autre symptôme du mouvement vers la Singularité : les idées vont se répandre de plus en plus vite, et même les courants les plus marginaux s'en trouveront normalisés. Quand j'ai commencé à écrire de la Science-Fiction au milieu des années 1960, il paraissait très facile de trouver des idées qui allaient mettre des décennies à tracer leur chemin dans la conscience culturelle ; aujourd'hui, j'évalue mon temps d'avance à environ dix-huit mois. (Bien sûr, c'est peut-être simplement l'imagination qui me fait défaut à mesure que je vieillis, mais je perçois cet effet chez d'autres auteurs.) Comme dans un choc mécanique, la Singularité se rapproche à mesure que nous tendons vers la vitesse critique.

Et que dire de l'arrivée de la Singularité ? De quoi cela aura-t-il l'air ? Puisqu'elle repose sur une fuite en avant intellectuelle, elle va probablement se produire plus vite que toutes les révolutions technologiques qui l'ont précédée. L'événement déclencheur sera vraisemblablement inattendu - y-compris pour les chercheurs impliqués. (*"Mais tous les modèles précédents étaient catatoniques ! Nous ne faisons qu'ajuster certains paramètres..."*) Si les technologies réseau sont suffisamment répandues (dans des

systèmes embarqués omniprésents), ces artefacts interconnectés pourraient donner aux utilisateurs l'impression d'être soudain *éveillés*.

Et que se passera-t-il un mois ou deux (ou un jour ou deux) après cela ? Je n'ai qu'un point de comparaison : l'avènement de l'Humanité. Nous serons passés dans l'ère post-Humaine. Et malgré mon optimisme technologique invétéré, je me dis parfois que je serais plus à mon aise à considérer ces événements transcendants à mille ans de distance... plutôt que vingt.

## La Singularité peut-elle être évitée ?

Bon, cela n'arrivera peut-être jamais. J'essaie parfois d'imaginer les symptômes à surveiller dans l'hypothèse où la Singularité ne viendrait pas à se développer. Il y a les arguments communément admis de Penrose [19] et Searle [22], qui questionnent la mise en oeuvre pratique de machines sapientes. En août 1992, Thinking Machines Corporation réunit un groupe de travail autour de la question : "*Comment construire une machine qui pense*" [27]. Comme l'indique l'intitulé de l'atelier, les participants n'étaient pas vraiment des détracteurs de l'intelligence des machines. En fait, il y avait un consensus autour du fait que l'intelligence peut exister dans des substrats non biologiques, et que les algorithmes sont des pré-requis à l'existence d'un esprit. Cependant, la question de la puissance matérielle brute présente dans les cerveaux organiques était loin de faire l'unanimité. Une minorité estimait que le rapport de puissance entre les ordinateurs les plus puissants en 1992 et le cerveau humain était de 1/3. La majorité des participants se ralliaient à l'estimation de Moravec [17] selon laquelle il faudrait attendre de 10 à 40 ans avant d'atteindre la parité homme/machine. Et un troisième groupe qui, se référant à des études sur le cerveau artificiel [7] [21], avançaient que les compétences computationnelles de neurones individuels sont en réalité bien plus développées qu'on ne le croit généralement. Selon eux, le matériel informatique d'aujourd'hui serait encore au bas mot 10 fois moins puissant que l'équipement que nous trébuchons dans nos têtes. S'ils ont raison (ou d'ailleurs, si les critiques de Penrose ou de Searle sont valides), la Singularité n'aura pas lieu. Au lieu de cela, en raison de notre incapacité à automatiser le travail de conception nécessaire à de nouvelles améliorations techniques, les performances du matériel cesseraient de progresser. Nous disposerions de matériel extrêmement puissant, mais demeurerions incapables d'aller plus loin. Les marchands nous proposeraient des systèmes de traitement numérique du signal toujours plus épatants, jusqu'à donner l'illusion de l'analogique à des opérations numériques, mais rien ne *s'éveillerait*, et on n'assisterait pas à cette fuite en avant intellectuelle qui est l'essence même de la Singularité. On peut voir cela comme un nouvel âge d'or, mais ce serait aussi la fin du progrès. Cela ressemble beaucoup au futur annoncé par Gunther Stent. De fait, Stent

explique que le développement d'une intelligence transhumaine\* suffit à réduire ses prédictions à néant. [25]

Toutefois, si la Singularité technologique peut survenir, elle le fera. Même si tous les gouvernements du monde devaient prendre la mesure de la *menace* et s'en méfier comme de la peste, le progrès continuerait de tendre vers ce point. La fiction évoque des lois passées pour prévenir la construction d'une "*machine qui ressemblerait à l'esprit humain*" [13]. Dans les faits, chaque avancée dans l'automatisation présente de tels avantages concurrentiels – économiques, militaires, voire artistiques – que passer des lois ou imposer des barrières douanières pour les interdire ne ferait que permettre à un autre d'y accéder le premier.

Eric Drexler [8] a proposé des réponses fabuleuses à la question : jusqu'où le progrès technique peut-il aller ? Il estime que des intelligences superhumaines seront accessibles dans un futur proche – et que de telles entités représentent une menace pour le status quo humain. Drexler estime cependant que l'on peut confiner ces machines superhumaines afin de les utiliser et de les étudier en toute sécurité. C'est la machine ultraintelligente selon I. J. Good, avec quelques précautions en plus. Pour moi, le confinement est intrinsèquement inenvisageable. Dans le cas d'un confinement physique : imaginons que l'on vous enferme dans votre maison, avec seulement un accès limité aux données provenant du dehors, de vos maîtres. Si ces maîtres pensent, disons, un million de fois plus lentement que vous, il est fort probable que vous finiriez, en quelques années de votre temps, par trouver une idée qui vous permettrait de vous libérer.

Je qualifie cette forme de superintelligence au *penser rapide* de *superhumanité faible*. Une telle entité, douée de *superhumanité faible*, se consumerait en quelques semaines du temps du dehors. La *superhumanité forte* ne se limite pas à accélérer l'horloge interne dans un système d'une puissance équivalente à l'esprit humain. Il est difficile de décrire ce que serait la *superhumanité forte*, mais la différence semble profonde. Imaginons que l'on accélère l'esprit d'un chien jusqu'à une très haute vitesse. Est-ce qu'un millier d'années de

---

\* La Transhumanité est le passage de l'Humanité à la post-Humanité (NdT)

pensée canine pourrait apporter quoi que ce soit à la perspicacité humaine ? (Quoique... Si l'esprit du chien était recâblé, puis lancé à haute vitesse, les résultats pourraient bien s'avérer différents.) Beaucoup de spéculations sur la superintelligence semblent basées sur le modèle de la superhumanité faible. Je crois que nos hypothèses seront mieux fondées en nous penchant vers la nature de la superhumanité forte. Je reviendrai plus loin sur ce point.

Une autre approche du confinement, c'est d'intégrer des règles dans l'esprit de notre création superhumaine (les lois de la robotique d'Asimov [3], par exemple). Je pense que toute règle suffisamment stricte pour être efficace amputerait du même coup l'entité d'une partie de ses capacités (et la compétition humaine favoriserait le développement de modèles plus dangereux). Néanmoins, le rêve d'Asimov est séduisant : imaginons un esclave volontaire qui serait 1000 fois meilleur que nous dans tous les domaines. Imaginons une créature qui pourrait satisfaire le moindre de nos désirs sains (quoi que cela signifie) et consacrer 99,9% de son temps à d'autres activités. Ce serait un nouvel univers que nous n'avons jamais vraiment compris, rempli de dieux bienveillants (au bout d'un moment, je pourrais même rêver de devenir l'un d'entre eux).

Si la Singularité ne peut être ni empêchée, ni confinée, est-ce que l'ère de la post-Humanité doit nécessairement s'avérer si horrible que ça ? Disons que... oui. L'extinction de la race humaine est une possibilité. (Ou, comme Eric Drexler le dit des nanotechnologies : vu tout ce dont ces technologies sont capables, il se pourrait que les gouvernements décident simplement qu'ils n'ont plus besoin de citoyens !) Pourtant, l'extinction n'est pas l'éventualité la plus effrayante. Encore une analogie : réfléchissons aux différents modes de relation que nous entretenons avec les animaux : élevage, dressage, animaux de compagnie, etc. Sans parler de la brutalité dont ils sont parfois victimes, le simple fait de les arracher à leur milieu naturel constitue un abus dont nous autres humains ne sommes pas toujours conscients. Dans le monde de la post-Humanité, il y aurait nombre de niches où des automates de niveau humain seraient souhaitables : des systèmes embarqués dans des appareils autonomes, des démons conscients en

charge des fonctions de bas niveau au sein d'êtres sensitifs plus vastes. (Une forme d'intelligence à superhumanité forte ressemblerait probablement à une Société d'Esprits [16] avec des composants très spécialisés.) Parmi les automates de niveau humain, certains pourraient n'être utilisés pour rien de plus que du traitement numérique du signal. Ils tiendraient plus des cétaqués que des humains. D'autres auraient une apparence humaine, mais avec un déséquilibre, une spécialisation obstinée qui, à notre époque, les conduirait à l'hôpital psychiatrique. Même si aucune de ces créatures n'est faite de chair et de sang, elles pourraient bien être, dans ce nouvel environnement, ce qui se rapproche le plus de ce que nous qualifions d'humain de nos jours. (I. J. Good s'est prononcé sur ce point, bien qu'à notre époque son argument ait l'air naïf. Good [12] propose une *Méta-Règle d'Or* que l'on peut exprimer ainsi : « *Traite tes inférieurs comme tu voudrais que tes supérieurs te traitent.* » C'est une idée merveilleuse et cependant paradoxale, parce qu'il est très difficile d'évaluer ce que nous aurions à y gagner. Pourtant, si on arrivait à s'y tenir, cela pourrait prouver que la bonté est possible dans cet univers.)

J'ai expliqué plus haut que nous ne pouvons empêcher la Singularité, que sa survenue est une conséquence inévitable de la compétitivité naturelle entre les humains et des possibilités qu'offre la technologie. Les plus grosses avalanches sont déclenchées par de tout petits riens. Et... c'est nous qui avons amorcé le mouvement. Cependant, bien qu'il soit difficile de déterminer la direction à suivre, nous avons la possibilité de poser les conditions initiales, de faire en sorte que les choses se passent le moins mal possible.

## D'autres chemins vers la Singularité : l'Intelligence Amplifiée

Quand on parle de créer des entités dotées d'une intelligence superhumaine, on pense généralement à des projets centrés sur l'IA. Pourtant, comme je l'ai précisé au début de cet article, il y a d'autres façons de parvenir à la superhumanité. Plutôt que les IA, ce sont les réseaux informatiques, les interfaces homme / machine qui, modestement, pourraient conduire à la Singularité. C'est ce que j'appelle l'*Intelligence Amplifiée*. La marche vers l'intelligence amplifiée est très naturelle. Le plus souvent, les développeurs eux-mêmes ne saisissent pas l'impact de leurs travaux. Mais chaque fois que nos capacités à accéder à l'information et à la partager augmentent, nous gravissons un échelon dans l'échelle de l'intelligence naturelle. Déjà, une équipe constituée d'un humain titulaire d'un doctorat et d'une bonne station de travail informatique (même hors ligne) pourrait atteindre le score maximal à n'importe quel test d'intelligence.

Et il semble plus aisé de parvenir à la superhumanité par le chemin de l'intelligence amplifiée que par celui de l'IA pure : plutôt que d'essayer de comprendre ce que nous sommes et de construire des machines à notre image, il paraît plus facile d'intégrer de nouvelles fonctionnalités, puisque les mystères du développement humain ont pour la plupart été élucidés. Quelques réflexions viennent étayer cette conjecture. Cairns-Smith [6] considère que la vie biologique pourrait s'être développée sur une forme de vie plus primitive basée sur la croissance cristalline. Lynn Margulis ([15] et ailleurs) a nettement explicité le rôle moteur du mutualisme dans l'évolution.

Notez bien que je ne propose pas d'ignorer les recherches sur l'IA, ou de lui couper les vivres. Les avancées en IA ont souvent des applications en intelligence amplifiée, et réciproquement. Je suggère simplement de reconnaître qu'il y a dans la recherche sur les réseaux et les interfaces quelque chose d'aussi profond (et potentiellement aussi incontrôlable) que dans celle sur l'IA. Des projets pourraient ainsi émerger sans avoir nécessairement d'application directe en terme d'interface ou d'architecture réseau

conventionnelles, simplement parce qu'ils permettent de progresser vers la Singularité sur le chemin de l'intelligence amplifiée.

Voici quelques-uns de ces projets :

- » L'automatisation du travail d'équipe humain / ordinateur. Il s'agit de prendre les problèmes pour lesquelles on opte généralement pour une solution informatique (comme les algorithmes d'optimisation), et de concevoir des programmes et des interfaces aptes à tirer parti à la fois des intuitions humaines et des capacités du matériel informatique. Confronté à la complexité un peu étrange d'un problème de *hill-climbing*<sup>\*</sup> de haut niveau, et puisqu'on dispose d'algorithmes ingénieux pour résoudre ce type de problèmes, le membre humain de l'équipe aurait à sa disposition des visualisations et des outils de contrôle très intéressants.
- » Développer une symbiose homme / machine dans la création artistique, en combinant les capacités de rendu graphique des machines modernes et le sens esthétique des humains. Bien sûr, un effort énorme a déjà été investi dans la conception d'outils informatiques d'aide à la création, pour faciliter le travail des artistes. Je suggère d'aller plus loin dans la fusion des compétences, de reconnaître explicitement la possibilité d'une approche véritablement coopérative. Les travaux de Karl Sims [23] sont autant d'avancées dans cette direction.
- » autoriser des équipes homme / machine dans les tournois d'échecs. Il existe déjà des programmes capables de battre quasiment n'importe quel humain. Et si cette compétence pouvait être exploitée par un humain, pour aller encore plus loin ? Ouvrir ne serait-ce que quelques tournois à des équipes mixtes pourrait avoir les mêmes effets positifs sur la recherche en intelligence amplifiée que l'ouverture des tournois d'échecs aux ordinateurs avait eu sur la

---

\* Le *hill-climbing* est une méthode de résolution de problèmes par approximations successives.  
Pour approfondir : [Wikipedia](#) (encore en anglais ; dès que j'ai 5 minutes, je m'y colle !) (NdT)

recherche en IA.

- » Développer des interfaces qui permettent aux humains d'accéder aux ordinateurs et aux réseaux sans être nécessairement assis devant un ordinateur. (Les avantages économiques que l'on pourrait retirer de cet aspect de l'intelligence amplifiée sont si évidents que des moyens considérables lui sont d'ores et déjà consacrés.)
- » Développer des systèmes d'aide à la prise de décision plus symétriques. L'informatique décisionnelle est d'ailleurs un secteur de recherche à la mode depuis quelques années. C'est une des composantes de l'intelligence amplifiée, qui me semble cependant encore trop centrée sur la machine. Il faut que l'utilisateur puisse guider le programme dans la même mesure que le programme lui fournit des informations.
- » Utiliser les réseaux locaux pour rendre les équipes humaines plus efficaces (c'est à dire plus efficaces que la somme des membres qui les composent). C'est ce qu'on appelle le *groupware*<sup>\*</sup>, un secteur déjà très prisé du fait de ses débouchés commerciaux. Il conviendrait toutefois changer de point de vue en considérant l'activité du groupe comme celle d'un organisme hybride. Dans un sens, on peut voir dans cette suggestion un appel à inventer un protocole pour régir ces organismes. Cela permettrait par exemple de focaliser la concentration du groupe plus facilement que dans les réunions classiques. Il deviendrait possible de se débarrasser des questions d'ego de façon à ce que l'expertise de chaque membre du groupe soit totalement mobilisée sur le projet. Et bien sûr, les bases de données pourraient être partagées de façon bien plus ergonomique que dans les comités habituels. (Cette suggestion vaut pour le travail d'équipe et non pour les meetings politiques : en effet, dans le champ politique, le protocole décrit plus haut ne ferait que renforcer le pouvoir de ceux qui font les règles !)

---

\* Le *groupware* regroupe l'ensemble des processus, protocoles et procédures, humains et informatiques, destinés à optimiser le travail de groupe. La traduction des mots anglais en *-ware* pose problème à cause du mot *logiciel* qui ne correspond pas à la réalité informatique, mais qui s'est cependant largement répandu. (NdT)

- » Exploiter Internet en tant qu'outil hybride homme / machine. De tous les points de cette liste, c'est le domaine qui évolue le plus vite et qui pourrait bien nous conduire le premier à la Singularité. Le pouvoir et l'influence de l'Internet d'aujourd'hui sont déjà largement sous-estimés. Je pense que nos systèmes informatiques actuels s'effondreraient sous le poids de leur propre complexité sans les coups de pouce que le "groupe de pensée" USENET donne à l'administration du système et aux supports techniques ! Le caractère anarchique du développement du réseau mondial est le meilleur témoin de son potentiel. La connectivité, la bande passante, les capacités de stockage et la vitesse des ordinateurs ne cessant d'augmenter, et nous voyons apparaître quelque chose qui ressemble à la vision de Lynn Margulis [15] : une biosphère capable de traiter de l'information, mais à des vitesses un million de fois plus grandes, et avec des millions d'agents à l'intelligence humaine (nous).

Les exemples ci-dessus illustrent des pistes de recherche que les départements de sciences informatiques peuvent commencer à explorer dès maintenant. Il y a d'ailleurs d'autres paradigmes. Par exemple, les travaux sur l'intelligence artificielle et les réseaux neuronaux gagneraient beaucoup à se rapprocher de la vie biologique. Plutôt que se limiter à essayer d'imiter la vie biologique avec des ordinateurs pour la comprendre, on devrait s'orienter vers la création de systèmes composites qui s'appuieraient sur la vie biologique soit comme simple support, soit pour fournir des fonctionnalités que nous ne comprenons pas encore suffisamment pour les implémenter au niveau matériel. C'est le vieux rêve de la Science-Fiction : l'interface directe cerveau / ordinateur [2] [29]. En fait, on peut travailler concrètement dans ce domaine – et on le fait déjà :

- » Les prothèses de membres offrent une application commerciale directe. On sait fabriquer des transducteurs entre les nerfs et le silicone. C'est une étape excitante, presque la dernière avant la communication directe.

- » Même si le débit reste faible, il semble faisable d'établir des liens directement à l'intérieur du cerveau. Grâce à la flexibilité d'apprentissage de l'humain, il n'est pas nécessaire de sélectionner trop précisément les neurones cibles dans le cerveau. Un débit, ne serait-ce que de 100 bits par seconde, pourrait être d'une grande utilité aux victimes d'attaques qui, sans cela, se retrouvent confinées à de simples interfaces avec menu.
- » En se branchant sur le nerf optique, il semble possible d'atteindre des débits approchant 1 Mbit/s. Mais pour cela, il nous faut connaître la micro-architecture de la vue, construire un énorme réseau d'électrodes avec une infinie précision. Si nous voulons que notre connection haut-débit s'ajoute aux chemins déjà présents dans le cerveau, le problème devient encore plus inextricable. Il ne suffit certainement pas de coller une grille de récepteurs haut-débit dans le cerveau. Mais supposons que la grille soit présente au moment où la structure cérébrale prend forme, pendant que l'embryon se développe. Ce qui implique :
- » Des expériences sur des embryons animaux. Je n'attends pas de véritables percées en IA dans les premières années de ces recherches, mais le développement de cerveaux en contact avec des structures neurales artificielles complexes pourrait intéresser des spécialistes du développement du cerveau embryonnaire. Sur le long terme, ces expérimentations pourraient produire des animaux avec des chemins sensitifs supplémentaires et des capacités cognitives intéressantes.

Au départ, je pensais que cette réflexion sur l'intelligence amplifiée pourrait mettre en lumière des chemins plus sûrs, moins dangereux, vers la Singularité. (Après tout, l'intelligence amplifiée ne nous invite-t-elle pas à prendre part à une sorte de transcendance ?) Hélas, en relisant ces propositions, tout ce dont je suis sûr c'est qu'il faut en tenir compte pour connaître toutes nos d'options. Mais quant à notre sauvegarde...

Certaines de ces suggestions sont suffisamment effrayantes sur le papier. Un de mes lecteurs m'a fait remarquer qu'en dotant quelques individus humains d'intelligence amplifiée, on crée une élite assez sinistre. Nous autres humains portons un bagage évolutionnaire de plusieurs millions d'années, un modèle où la compétition pour la survie est une lutte à mort. Cette violence n'est peut-être plus nécessaire dans le monde d'aujourd'hui, où les perdants imitent les attitudes et les trucs des gagnants pour se faire coopter dans leurs entreprises. Mais une créature construite à partir de rien ne serait-elle pas une entité beaucoup plus bénigne qu'une créature dont le *kernel*<sup>\*</sup> repose sur la loi du Talion ? Même la perspective d'un Internet égalitaire qui s'éveillerait en même temps que l'Humanité tout entière ressemble fort à un cauchemar [26].

Le problème ne se limite pas au fait que la Singularité correspond à la sortie de scène de l'Humanité. En effet, ses implications se heurtent à nos plus profondes représentations de l'Être. C'est ce qu'une réflexion plus fine sur la superhumanité forte peut démontrer.

---

\* Informatique : le *kernel* est le cœur du système ; il fournit aux logiciels (*software*) des points d'accès au matériel (*hardware*). (NdT)

## La superhumanité forte et ce que nous pouvons attendre de mieux.

Supposons que nous puissions évoluer pour nous adapter à la Singularité. Supposons que nous puissions réaliser nos espoirs les plus extravagants. Que nous resterait-il à attendre ? Que les humains deviennent leurs propres successeurs, que la connaissance de nos origines nous permette de faire face à n'importe quelle injustice. Ce serait un âge d'or où le progrès resterait possible (s'affranchissant au passage de la limite posée par Stent). Chacun pourrait prétendre à l'immortalité (où du moins à une vie qui durerait aussi longtemps que nous sommes capables de faire survivre l'univers [10] [4]).

Mais dans ce monde si brillant et si bon, ce sont les problèmes philosophiques qui deviennent menaçants. Un esprit qui reste au même niveau de capacité ne peut vivre indéfiniment ; après quelques milliers d'années, il aurait plus l'air d'une cassette qui tourne en boucle que d'une personne. (C'est dans [18] que j'ai vu l'illustration la plus effrayante de ce point.) Pour vivre indéfiniment, l'esprit lui-même doit pouvoir grandir... Et quand il est devenu suffisamment grand et qu'il regarde en arrière, quelles affinités lui reste-t-il avec l'esprit qu'il était au départ ? Nul doute que le nouvel être est en tout point identique à l'original, mais bien plus intensément. Ainsi, même à l'échelle de l'individu, la notion d'une nouvelle vie qui se développe sur une ancienne, comme Cairns-Smith ou Lynn Margulis l'ont proposée, semble se vérifier.

Le *problème* de l'immortalité surgit plus manifestement ailleurs. Les notions d'ego et de conscience de soi sont les piliers de la pensée rationaliste pure et dure de ces deux ou trois derniers siècles. Or la notion de conscience de soi subit aujourd'hui les attaques des chercheurs en Intelligence Artificielle ("*conscience de soi et autres illusions*"). Et c'est l'Intelligence Amplifiée qui sape nos conceptions quant à la notion d'ego. En effet, le monde de la post-Singularité reposera sur des liaisons réseau à très haut débit. Et la capacité à communiquer à des débits variables, parfois bien plus élevés que ceux de la communication parlée ou écrite, sera déterminante pour que des entités à superhumanité

forte puissent copier, échanger, fusionner des morceaux d'ego, agrandir ou réduire à volonté la taille de leur conscience de soi afin de s'adapter aux problèmes à traiter. Ce sont des caractéristiques essentielles de la superhumanité forte et de la Singularité.

À bien y penser, on réalise que l'ère post-Humaine sera profondément étrange et très différente de la nôtre – même dans l'hypothèse où l'on ferait tout pour s'y préparer intelligemment et avec bienveillance.

D'un certain point de vue, cette vision satisfait à nos plus beaux rêves : un temps infini, qui permet à tous de se connaître vraiment et de comprendre les mystères les plus impénétrables. Vu sous un autre angle, cela ressemble beaucoup au scénario catastrophe que j'ai exposé plus haut.

Quelle est la meilleure option ? En fait, la nouvelle ère est trop différente pour rentrer dans le cadre traditionnel du bien et du mal. Ce schéma convient pour des esprits isolés, immuables, connectés par des liens ténus à bas-débit. Cependant, le monde de la post-Singularité s'inscrit dans une tradition, celle du changement et de la coopération, bien antérieure au manichéisme (et peut-être même à l'émergence de la vie biologique). Je pense que cette ère aura toutefois besoin de quelques notions d'éthique. C'est ce que les chercheurs en Intelligence Amplifiée et en communications haut débit doivent comprendre. Je n'en perçois moi-même pas encore toutes les implications [32]. En s'inspirant de la Méta-Règle d'Or de Good, on doit pouvoir établir des règles pour différencier Soi et Autrui, basées sur le débit de la connection. L'esprit et le Soi seront incomparablement plus labiles que par le passé, mais nous ne devrions pas perdre ce qui importe vraiment : la connaissance, la pensée, le souvenir. Je pense que Freeman Dyson a raison de dire : *"Dieu est ce qu'il advient de l'esprit une fois qu'il a franchi les limites de notre entendement."*

[J'aimerais remercier John Carroll de San Diego State University et Howard Davidson de Sun Microsystems d'avoir bien voulu discuter avec moi des ébauches de cet article.]

## Bibliographie annotée

(et, au passage, un appel à candidats pour le travail bibliographique...)

- [1] Alfvén, Hannes, sous le pseudonyme de Olof Johanneson, *The End of Man ?* Award Books, 1969 ; publié auparavant sous le titre "The Tale of the Big Computer", Coward-McCann, copyright 1966 Albert Bonniers Forlag AB – pour la traduction anglaise : copyright 1966 Victor Gollanz, Ltd.
- [2] Anderson, Poul, "Kings Who Die", *If*, March 1962, p8-36. Republié dans *Seven Conquests*, Poul Anderson, MacMillan Co., 1969.
- [3] Asimov, Isaac, "Runaround", *Astounding Science Fiction*, Mars 1942, p94. Republié dans *Robot Visions*, Isaac Asimov, ROC, 1990. Asimov raconte dans ce livre la maturation de ses histoires de robots.
- [4] Barrow, John D. et Frank J. Tipler, *The Anthropic Cosmological Principle*, Oxford University Press, 1986.
- [5] Bear, Greg, "Blood Music", *Analog Science Fiction-Science Fact*, Juin, 1983. Développé dans le roman *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, pp175-218, 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, pp447-460, 1979.
- [11] Good, I. J., "Speculations Concerning the First Ultra-intelligent Machine", in *Advances in Computers*, vol 6, Franz L. Alt et Morris Rubinoﬀ, eds, pp31-88, 1965, Academic Press.
- [12] Good, I. J., [Au secours! Je ne retrouve plus la source de la Méta-Règle d'Or de Good. Je me souviens nettement en avoir entendu parler dans les années 1960. Grâce au net, j'ai déniché quelques liens vers des pages apparentées. G. Harry Stine et Andrew Haley parlent d'une méta-loi, appliquée aux extraterrestres : G. Harry Stine, "How to Get along with Extraterrestrials... or Your Neighbor", *Analog Science Fiction-Science Fact*, Février, 1980, p39-47.]
- [13] Herbert, Frank, *Dune*, Berkley Books, 1985. Ce roman a paru sous forme de feuilleton dans *Analog Science Fiction-Science Fact* dès les années 1960s.
- [14] Kovacs, G. T. A. et al., "Regeneration Microelectrode Array for Peripheral Nerve Recording and Stimulation", *IEEE Transactions on Biomedical Engineering*, v 39, n 9, pp 893-902.
- [15] Margulis, Lynn et 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*, Avril 1967, pp82-108. Republié dans *Neutron Star*, Larry Niven, Ballantine Books, 1968.
- [19] Penrose, Roger, *The Emperor's New Mind*, Oxford University Press, 1989.
- [20] Platt, Charles, communication privée.
- [21] Rasmussen, S. *et al.*, "Computational Connectionism within Neurons: a Model of Cytoskeletal Automata Subservicing Neural Networks", in *Emergent Computation*, Stephanie Forrest, ed., pp428-449, MIT Press, 1991.
- [22] Searle, John R., "Minds, Brains, and Programs", in *The Behavioral and Brain Sciences*, vol 3, Cambridge University Press, 1980. L'essai paraît aussi dans *The Mind's I*, édité par Douglas R. Hofstadter et Daniel C. Dennett, Basic Books, 1981 (ma source pour cette référence). Cette réédition contient une excellente critique de l'essai de Searle.
- [23] Sims, Karl, "Interactive Evolution of Dynamical Systems", Thinking Machines Corporation, Technical Report Series, publié dans *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, Paris, MIT Press, Décembre 1991.
- [24] Stapledon, Olaf, *The Starmaker*, Berkley Books, 1961 (probablement commencé avant 1937).
- [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*, paru sous forme de feuilleton dans *Isaac Asimov's Science Fiction Magazine*, Décembre(?) 1986 - Février 1987. Republié par Ace Books, 1988.
- [27] Thearling, Kurt, "*How We Will Build a Machine that Thinks*", un atelier organisé par Thinking Machines Corporation, 24-26 Août 1992. Communication privée.
- [28] Ulam, S., Tribute to John von Neumann, *Bulletin of the American Mathematical Society*, vol 64, nr 3, part 2, Mai 1958, pp1-49.
- [29] Vinge, Vernor, "Bookworm, Run!", *Analog*, Mars 1966, pp8-40. Republié dans *True Names and Other Dangers*, Vernor Vinge, Baen Books, 1987.
- [30] Vinge, Vernor, "True Names", *Binary Star Number 5*, Dell, 1981. Republié dans *True Names and Other Dangers*, Vernor Vinge, Baen Books, 1987.
- [31] Vinge, Vernor, "First Word", *Omni*, January 1983, p10.
- [32] Vinge, Vernor, À paraître :-)