

Gâchis Astronomique: le coût d'opportunité des délais en développement technologique

Par Nick Bostrom

Abstract: Grâce à des technologies avancées, on pourrait maintenir une très grande quantité de personnes menant des vies heureuses dans la région accessible de l'univers. Chaque année où la colonisation de l'univers ne se déroule pas représente un coût d'opportunité; des vies qui valent d'être vécues ne peuvent être réalisées. D'après des estimations plausibles, ce coût est extrêmement élevé. Mais la leçon pour les utilitaristes n'est pas qu'il faut maximiser la *cadence* du développement technologique, mais sa *sécurité*. Autrement dit, il faut maximiser la probabilité que la colonisation se déroule.

I. Le rythme de perte de vies potentielles

En ce moment, des soleils illuminent et réchauffent des pièces vides et des trous noirs absorbent une portion de l'énergie inutilisée du cosmos. Chaque minute, notre héritage de néguentropie se dissipe irréversiblement en entropie. Ces ressources ainsi gâchées auraient pu être utilisées par une civilisation avancée au bénéfice d'êtres sensibles menant de bonnes vies.

Cette perte se fait à un rythme ahurissant. Un papier récent postule par des considérations théoriques librement basées sur le rythme d'accroissement de l'entropie, que la perte de vies humaines potentielles dans notre superamas galactique est d'au moins $\sim 10^{46}$ vies par siècle de colonisation retardée. [1] Cette estimation suppose que toute l'entropie perdue aurait pu être utilisée à des fins productives. Mais puisqu'aucun mécanisme technologique n'est encore capable de faire cela, ce n'est pas une estimation suffisamment prudente.

Nous pouvons obtenir notre estimation de la limite inférieure en multipliant le nombre d'étoiles de notre superamas galactique par la puissance de calcul qu'on peut extraire par étoile, le tout divisé par la puissance requise pour simuler une vie humaine.

À titre de grossière approximation, disons que le Superamas de la Vierge est composée de 10^{13} étoiles. On estime la puissance de calcul qu'on peut extraire d'une étoile en utilisant de la nanotechnologie moléculaire avancée [2] à 10^{42} opérations par seconde [3]. L'estimation typique de la puissance de calcul du cerveau humain est un peu moins de 10^{17} opérations par seconde [4]. Notons qu'il ne semble pas nécessiter davantage de puissance de calcul en plus

pour simuler tous les détails pertinents de l'environnement d'un humain typique [5]. Cela implique qu'un potentiel d'environ 10^{38} vies humaines est perdu chaque siècle, ce qui revient à 10^{29} vies par seconde.

Cette estimation est prudente dans la mesure où elle n'envisage que des techniques de calcul déjà esquissées dans la littérature scientifique. On peut obtenir une estimation encore moins ambitieuse en supposant que les humains du futur s'incarnent uniquement dans des corps biologiques. Supposons qu'environ 10^{10} humains biologiques puissent vivre autour d'une étoile moyenne. Le Superamas de la Vierge pourrait alors contenir 10^{23} humains biologiques. D'après cette estimation, la perte de potentiel est donc égale à 10^{14} vies humaines perdues par seconde de délai.

Ce ne sont pas les chiffres exacts qui nous importent pour le moment, mais le fait qu'ils soient astronomiques. Même notre estimation la plus prudente (qui suppose une incarnation biologique des personnes potentielles) implique un rythme de cent mille milliards de vies humaines potentielles perdues par seconde de colonisation retardée [6].

II. Le coût d'opportunité des retards de la colonisation

D'un point de vue utilitariste, cette perte de vies humaines correspond à une perte de valeur potentielle. Il est communément admis que les vies humaines actuelles valent d'être vécues. Une civilisation assez avancée pour coloniser le superamas local pourra sans doute établir les conditions minimales pour dire de même.

Ainsi, accélérer le développement technologique (ou ses facteurs déterminants, comme la productivité économique) semble avoir plus d'effet sur la valeur totale que n'importe quelle autre action. Agir de telle sorte à ce que la colonisation se déroule juste *une seconde* plus tôt qu'elle ne se déroulerait normalement revient à engendrer 10^{29} vies humaines (10^{14} vies humaines avec l'estimation la plus conservatrice). Peu de causes philanthropiques peuvent rivaliser avec ce niveau de rendement utilitaire.

D'autres points de vue que l'utilitarisme sur ce qui constitue la valeur arrivent à la même conclusion. Par exemple, une conception du bien-être humain qui va plus loin que l'utilitarisme typique pourrait mesurer sa valeur dans l'épanouissement et l'expression de l'individu, les relations significatives, et la noblesse de caractère: cela importe peu. Tant que la mesure de valeur est agrégative (le bonheur de l'un ne fait pas le malheur de l'autre) et est sans actualisation temporelle, cette conclusion tiendra.

Elle pourra même toujours tenir si la mesure de valeur n'est pas parfaitement agrégative (par exemple si un aspect de la valeur est la diversité, dont la production marginale risque de décliner avec une population grandissante) tant qu'un composant important de la valeur reste agrégatif. De même, un degré minimal d'actualisation temporelle peut être adopté sans changer de conclusion [7].

III. Le grand but des utilitaristes devrait être de réduire le risque existentiel

On pourrait penser conclure ainsi, avec la résolution qu'un utilitariste devrait logiquement concentrer ses efforts sur l'accélération du développement technologique. Après tout, les bénéfices d'une réussite infime dans ce projet éclipsent presque toute autre activité. Nous nous trouvons en possession d'un argument utilitariste prônant une urgence absolue dans le développement technologique.

Mais la véritable leçon est toute autre. Il ne suffit pas de prendre en compte le coût d'opportunité des délais de colonisation, mais aussi le risque d'échec. Nous pourrions être affligés par un *risque existentiel*, une catastrophe qui éliminerait toute vie intelligente sur Terre ou amputerait autrement son futur. La longévité des galaxies se mesure dans les milliards d'années, tandis que tout délai de colonisation ne se mesurerait qu'en années ou décennies. Ainsi, la considération des risques triomphe largement sur la considération du coût d'opportunité. Une réduction des risques existentiels d'un seul pourcent correspond (d'un point de vue d'utilité espérée) à un délai de plus de 10 millions d'années.

Ainsi, l'effet infime d'une action sur la *probabilité* d'une colonisation éventuelle l'emporte sur l'effet concernant sa vitesse. Pour les utilitaristes typiques, la priorité une, deux, trois, et quatre devrait être de réduire le risque existentiel. Le cri utilitariste "Maximisons l'utilité agrégative espérée!" peut être simplifiée par la maxime "Minimisons le risque existentiel!"

IV. Implications pour les vues effets-sur-personnes agrégatives

L'argument ci-dessus suppose que notre but est de maximiser la quantité totale de bien-être. Supposons plutôt que nous adoptons la version "effets-sur-personnes" de l'utilitarisme, d'après qui nos obligations sont envers les personnes qui existent à l'instant [9]. Ainsi, l'extinction humaine serait mauvaise parce qu'elle rendrait pire les vies présentes, et non parce qu'elle constitue une perte de vie potentielles. Que devrait faire une personne qui suit cette doctrine? Devrait-elle mettre l'emphase sur la vitesse, la sécurité... quelque chose d'autre encore?

Pour répondre, il nous faut considérer d'autres variables. Supposons qu'un tel pense que la probabilité qu'une personne vivante à l'instant puisse bénéficier des ressources extraites de la colonisation est faible. Sa raison de s'opposer aux risques existentiels serait donc que l'extinction humaine amputerait (disons) une moyenne de 40 ans à la vie de 8 milliards de vies humaines [10]. Cela serait certainement une catastrophe; mais elle serait du même ordre de grandeur que d'autres tragédies humaines comme la maladie ou la faim. Un utilitariste effets-sur-personnes devrait ainsi voir la réduction de risques existentiels comme étant une priorité importante, mais pas absolument dominante. Dans ce cas, il n'est pas immédiatement évident ce qu'il doit faire. La réponse à cette question dépendra de calculs détaillés pour trouver quelle cause philanthropique se trouve le mieux à sa portée.

Mais il est possible qu'il lui soit nécessaire en tout cas de donner une probabilité non négligeable à la proposition que des personnes toujours vivants aujourd'hui puissent survivre assez longtemps pour accéder aux ressources astronomiques. Il se pourrait qu'une "singularité technologique" se déroule durant sa vie naturelle [11], ou une révolution dans l'extension biologique de la vie, ou même dans la nanotechnologie; dans tous les cas, le processus de vieillesse pourrait être effacé et inversé [12]. Beaucoup de scientifiques et de futurologues donnent une haute probabilité que ces technologies se développent dans les prochaines décennies [13]. Si ces pronostics vous semblent irréalistes, considérez la performance abyssale des prévisions technologiques passées. Ce serait un excès de confiance que de donner une probabilité de moins (disons) d'1% à la proposition que l'on verra ses technologies se développer durant notre vie.

Un nombre astronomique divisé par cent est toujours un nombre astronomique: l'utilité espérée que représente 1% de chance que ces technologies de réalisent est immense. Ainsi, quelle est l'utilité espérée dans le cas où une portion significative de la population accède aux ressources astronomiques? La réponse n'est pas donnée. D'un côté, on peut noter que l'utilité marginale des ressources matérielles diminue fortement après un certain point. Après tout, le niveau de bien-être de Bill Gates ne semble pas surpasser dramatiquement celui d'une personne aux moyens plus modestes. D'un autre côté, des technologies avancées—qui seront sans doute déployées à l'époque de la colonisation du superamas—pourraient nous offrir de nouveaux moyens de convertir des ressources matérielles en bien-être. On pourrait par exemple utiliser des ressources pour augmenter nos capacités mentales et prolonger indéfiniment notre espérance de vie subjective. Il n'est pas du tout donné que l'utilité marginale d'une augmentation de l'esprit et de la longévité doit fortement diminuer après un certain point. Ainsi, l'utilité espérée que représente pour une personne vivante aujourd'hui l'accès aux ressources issues de la colonisation du superamas est *astronomique*. Cette conclusion tient *même si* on donne une faible probabilité à ces suppositions. Pour un maximiseur de l'utilité espérée, le bien-être apporté par des milliards d'années de vie subjectives et une augmentation des capacités mentales est immense, même si la probabilité de réussite est faible.

Alors, enfin: que devrait faire un utilitaire effets-sur-personnes? Il est clair qu'il est important d'éviter les catastrophes existentielles, non seulement parce qu'elles amputerait l'espérance de vie de 8 milliards de personnes, mais aussi (et nos suppositions démontrent que c'est une considération plus imposante) parce qu'elles anéantiraient tout espoir à ces personnes d'accéder aux ressources astronomiques. L'utilitariste total peut se concentrer entièrement sur les risques existentiels. Mais pour une utilitariste effets-sur-personnes, il ne suffit pas que l'humanité survive pour coloniser le superamas: il faut aussi que les personnes actuelles survivent *personnellement*. Elle doit donc mettre l'accent sur la rapidité des développements technologiques, afin de permettre aux personnes actuelles de survivre jusqu'à ce que les fruits de la colonisation soient récoltés (ce qui implique des technologies d'extension de la longévité). Si la vitesse entre en conflit avec la sécurité, l'utilitariste total doit toujours prioriser la sécurité. L'utilitariste effets-sur-personne, elle, doit soigneusement faire l'équilibre entre le risque que tout le monde meurt de vieillesse et celui que tout le monde meurt d'une catastrophe d'extinction humaine [14].

[1] M. Cirkovic, 'Cosmological Forecast and its Practical Significance', *Journal of Evolution and Technology*, xii (2002), <https://www.jetpress.org/volume12/CosmologicalForecast.pdf>.

[2] K. E. Drexler, *Nanosystems: Molecular Machinery, Manufacturing, and Computation*, New York, John Wiley & Sons, Inc., 1992.

[3] R. J. Bradbury, 'Matrioshka Brains', Manuscript, 2002, <http://www.aeiveos.com/~bradbury/MatrioshkaBrains/MatrioshkaBrains.html>

[4] N. Bostrom, 'How Long Before Superintelligence?', *International Journal of Futures Studies* ii (1998); R. Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, New York, Viking, 1999. L'estimation la plus basse est dans H. Moravec, *Robot: Mere Machine to Transcendent Mind*, Oxford, 1999.

[5] N. Bostrom, 'Are You Living in a Simulation?', *Philosophical Quarterly*, liii (211). Voir aussi <https://simulation-argument.com>.

[6] Le superamas de la Vierge ne représente qu'une infime partie des ressources colonisables de l'univers, mais est de taille suffisamment astronomique pour étayer nos propos. Agrandir notre région d'intérêt rendrait exponentiellement plus incertain la proposition qu'on soit la seule civilisation intelligente à profiter des ressources astronomiques.

[7] Il est admis par les utilitaristes que l'actualisation temporelle n'est pas approprié pour évaluer la valeur (voir e.g. R. B. Brandt, *Morality, Utilitarianism, and Rights*, Cambridge, 1992, pp. 23f.). Cependant, les utilitaristes pourraient être contraints à compromettre ce principe dans la mesure où nos actions pourraient avoir des conséquences qui touchent un nombre infini de personnes (une possibilité que nous mettons pour l'instant de côté).

[8] N. Bostrom, 'Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards', *Journal of Evolution and Technology*, ix (2002), <https://www.jetpress.org/volume9/risks.html>.

[9] Cette formulation n'est pas nécessairement la meilleure pour décrire cette position, mais elle est simple, et suffira pour nos propos.

[10] Ou ce que sera la population humaine juste avant l'apocalypse.

[11] Voir e.g. V. Vinge 'The Coming Technological Singularity', *Whole Earth Review*, Winter issue (1993).

[12] R. A. Freitas Jr., *Nanomedicine*, Vol. 1, Georgetown, Landes Bioscience, 1999.

[13] E.g. Moravec, Kurzweil, and Vinge op. cit.; E. Drexler, *Engines of Creation*, New York, Anchor Books, 1986.

[14] Je suis reconnaissant du support financier offert par le British Academy Postdoctoral Award.