STIAMO VIVENDO IN UNA REALTÀ SIMULATA?

di Nick Bostrom

traduzione di claudiordali

[Pubblicato su *Philosophical Quarterly* (2003)Vol.53 Nr. 211 Pagg. 243-255 (Prima Versione del 2001)]

Versione originale in inglese

Questo scritto sostiene che *almeno* una delle seguenti affermazioni è vera: (1) la specie umana è molto probabile che si estingua prima di raggiungere una fase "post-umana"; (2) ogni civiltà post-umana è estremamente difficile che produca un numero significativo di simulazioni della propria storia evolutiva (o variazioni della stessa); (3) stiamo quasi certamente vivendo in una realtà simulata. Ne consegue che la convinzione che ci sia una possibilità significativa per cui un giorno diventeremo post-umani e produrremo simulazioni ancestrali, sia falsa, a meno che non stiamo già vivendo in una simulazione. Verranno discusse un certo numero di altre conseguenze di questo risultato.

I. INTRODUZIONE

Molte opere di fantascienza, nonché alcune previsioni fatte da tecnologi e futurologi seri, prevedono che in futuro saranno disponibili enormi quantità di energia per l'elaborazione dati. Supponiamo per un momento che queste previsioni siano corrette. Una cosa che le generazioni successive potranno fare con i loro computer super-potenti, sarà quella di produrre in modo dettagliato simulazioni ancestrali tipo quelle dei loro progenitori o individui simili. Siccome i loro computer saranno molto potenti, potranno eseguire moltissime di queste simulazioni. Supponiamo che questi individui simulati siano consapevoli (e potrebbero esserlo, se le simulazioni saranno sufficientemente elaborate nei dettagli e se una specifica posizione, ampiamente accettata nella filosofia della mente, sarà corretta). Allora, potrebbe essere il caso che la stragrande maggioranza delle menti come la nostra, non appartenga alla razza originale, ma piuttosto agli individui simulati per mezzo dei discendenti avanzati di una razza originale. Se fosse il caso, è quindi possibile sostenere che sarebbe razionale pensare che le menti simulate abbiano delle probabilità di successo piuttosto che quelle biologiche originali. Pertanto, se non pensiamo di vivere realmente in una simulazione della realtà, non abbiamo il diritto di credere che avremo dei discendenti che produrranno un sacco di simulazioni ancestrali. Questa è l'idea di base. Il resto di questo scritto la descriverà con molta attenzione.

A parte l'interesse per questa tesi da parte di coloro che sono stanno lavorando a una speculazione futuristica, ci sono anche dei bonus più puramente teorici. L'argomento fornisce lo stimolo per formulare qualche questione metodologica e metafisica e suggerisce analogie naturalistiche per alcune concezioni religiose tradizionali, che alcuni potrebbero trovare divertenti o provocanti.

La struttura dello scritto è la seguente. In primo luogo, formuliamo l'ipotesi che, al fine di iniziare l'argomentazione, c'è bisogno di importare del materiale dalla filosofia della mente. In secondo luogo, consideriamo alcune ragioni empiriche, in modo da poter pensare che la produzione di un gran numero di simulazioni della mente umana potrebbe essere alla portata di una civiltà futura, qualora avesse già sviluppato molte di quelle tecnologie che si sono dimostrate compatibili con le leggi fisiche ei vincoli ingegneristici conosciuti. Questa parte non è filosoficamente necessaria, ma fornisce un incentivo per prestare attenzione al resto. Segue poi il nucleo dell'argomento, che fa uso di qualche semplice teoria delle probabilità e un paragrafo che fornisce un sostegno al blando principio di indifferenza utilizzato dall'argomento. Infine, saranno discusse alcune interpretazioni della disgiunzione, citate in astratto, che costituiscono la conclusione della discussione sulla simulazione.

II. LA PREMESSA PER UN SUBSTRATO INDIPENDENTE

Un presupposto comune nella filosofia della mente, è quello di un *Substrato Indipendente.* L'idea è che gli stati mentali possono sopravvenire a qualsiasi ampia categoria di substrati fisici. Una volta fornite di un sistema che implementa i giusti tipi di strutture e processi computazionali, potranno essere associate ad esperienze coscienti. Non è una proprietà essenziale della coscienza, quella di essere implementata su reti neurali biologiche a base di carbonio all'interno di un cranio: in linea di principio, potrebbero essere validi anche dei processori a base di silicio posti all'interno di un computer. Gli argomenti per questa tesi sono stati presi dalla letteratura e sebbene siano abbastanza controversi, in questo ambito verranno presi come dati.

L'argomento che presenteremo, tuttavia non dipende da alcuna versione molto considerevole di funzionalismo o computazionalismo. Ad esempio, non abbiamo bisogno di dare per scontato che la tesi del substrato indipendente sia *necessariamente* vera (sia analiticamente che metafisicamente) - diciamo solo che di fatto, un computer con un idoneo programma, potrebbe essere cosciente. Inoltre, non dobbiamo supporre che per creare una mente in un computer sarebbe sufficiente programmarlo in modo tale che si comporti come un essere umano in qualsiasi situazione, compreso superare il test di Turing ecc.. Abbiamo bisogno solo che, alla generazione di

esperienze soggettive, sarebbe sufficiente l'ipotesi più debole per cui i processi di calcolo di un cervello umano vengano strutturalmente replicati con dettagli opportunamente fini, tipo il livello delle 3 sinapsi dell'individuo. Questa versione attenuata del substrato indipendente è abbastanza ampiamente accettata.

I neurotrasmettitori, i fattori di crescita dei nervi e le altre sostanze chimiche che sono più piccole di una sinapsi, giocano chiaramente un ruolo importante nell'apprendimento e nella cognizione umana. La tesi del substrato indipendente non vuol dire che gli effetti di queste sostanze chimiche sono minimi o irrilevanti, ma piuttosto che interessano l'esperienza soggettiva solo attraverso la loro diretta o indiretta influenza sulle attività computazionali. Ad esempio, se non ci può essere differenza nell'esperienza soggettiva, senza che vi sia anche una differenza nelle emissioni sinaptiche, allora la precisione necessaria per la simulazione è a livello sinaptico (o superiore).

III. I LIMITI TECNOLOGICI DELLA COMPUTAZIONE

Al nostro attuale stadio di sviluppo tecnologico, non abbiamo né un hardware sufficientemente potente, né il software necessario per creare una mente cosciente in un computer. Ma gli argomenti persuasivi sono stati trattati per il fatto che *se* il progresso tecnologico non accenna a diminuire, *allora* queste carenze saranno alla fine superate. Alcuni autori sostengono che questa fase può essere superata in alcuni decenni.[1] Gli scopi attuali non richiedono ancora ipotesi circa i tempi. L'argomento della simulazione funziona altrettanto bene per chi pensa che ci vorranno centinaia di migliaia di anni prima di raggiungere la fase "post-umana" della civiltà, la fase in cui l'umanità avrà acquisito la maggior parte della capacità tecnologica che si può attualmente dimostrare essere coerente con le leggi fisiche e con i vincoli della materia e dell'energia.

Una tale fase matura di sviluppo tecnologico renderà possibile convertire i pianeti e le altre risorse astronomiche in computer enormemente potenti. Al momento è difficile essere sicuri su qualsiasi limite massimo riguardo alla potenza di elaborazione dati che potrà essere a disposizione della civiltà post-umana. Siccome siamo ancora privi di una "teoria del tutto", non possiamo escludere la possibilità che nuovi fenomeni fisici, non ammessi nelle teorie fisiche attuali, possano venire utilizzati per superare quei vincoli[2], che nella nostra attuale comprensione, impongono limiti teorici sull'elaborazione delle informazioni ottenibili in una determinata massa di materia.

---

[1] Vedere e.g K.E. Drexler, *Engines of Creation: The Coming Era of Nanotechnology,* Londra, 4th Estate, 1985; N. Bostrom, "How Long Before Superintelligence?" *International Journal of Futures Studies,* vol.2, (1988); R. Kurzweil, *The Age of Spiritual Machines: When computers exceed human intelligence,* New York, Viking Press, 1999; H. Moravec, *Robot: Mere Machine to Transcendent Mind*, Oxford University Press, 1999. [2] Tipo il vincolo di Bremermann-Bekenstein e il limite del buco nero (H.J. Bremermann, "Minimum

Energy requirements of information transfer and computing." *International Journal of Theoretical Physics* 21: 203-217 (1982); J. D. Bekenstein, "Entropy content and information flow in systems with limited energy." *Physical Review* D 30: 1669-1679 (1984); A. Sandberg, "The Physics

Possiamo stabilire, con molta più fiducia, limiti inferiori alla computazione post-umana, ipotizzando solo i meccanismi che sono già stati compresi. Ad esempio, Eric Drexler ha delineato un progetto per un sistema che taglia una zolletta di zucchero (ad esclusione del raffreddamento e dell'alimentazione) che svolgerà 1021 istruzioni al secondo.[3] Un altro autore dà una stima approssimativa di 1042 operazioni al secondo per un computer con una massa nell'ordine di un grande pianeta.[4] (Se potessimo creare computer quantici, o imparare a costruire computer dalla materia nucleare o dal plasma, potremmo spingerci più vicini ai limiti teorici. Seth Lloyd calcola il limite superiore per un computer di 1 kg, pari a 5*1050 operazioni logiche svolte al secondo su ~ 1031 bit.[5] Tuttavia, per i nostri scopi, è sufficiente utilizzare una stima più conservativa che presupponga solo design e principi attualmente conosciuti.)

La quantità di potenza di calcolo necessaria per emulare una mente umana può essere approssimativamente stimata nello stesso modo. Secondo una stima, basata su quanto sia stato computazionalmente costoso replicare la funzionalità di un pezzo di tessuto nervoso con del silicio, il miglioramento del contrasto nella retina produce un numero di ~ 1014 operazioni al secondo per l'intero cervello umano.[6] Una stima alternativa, basata sul numero di sinapsi nel cervello e sulla loro frequenza di emissione, produce una cifra complessiva di ~ 1016-1017 operazioni al secondo.[7] In teoria, potrebbero esserne necessarie ancora di più, se vogliamo simulare nei dettagli il funzionamento interno delle sinapsi o degli alberi dendritici. Tuttavia, è probabile che il sistema nervoso centrale umano abbia un alto grado di ridondanza su microscale, per compensare l'inaffidabilità e la rumorosità dei suoi componenti neuronali. Ci si aspetterebbe quindi un notevole guadagno in efficienza dall'utilizzo di processori non biologici più affidabili e versatili. Sembra che la memoria non sia più un vincolo rigoroso nella potenza di elaborazione.[8] Inoltre, dato che la larghezza massima della banda sensoriale umana è di ~ 108 bit al secondo, simulare tutti gli eventi sensoriali comporta un costo trascurabile rispetto a simulare l'attività corticale.

---

of Information Processing Superobjects: The Daily Life among the Jupiter Brain." *Journal of Evolution and Technology*, vol. 5 (1999).

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

[4] R. J. Bradbury, "Matrioshka Brains." Working Manuscript (2002), http://www.aeiveos.com/~bradbury/MatrioshkaBrains/MatrioshkaBrains.html.

[5] S. Lloyd, "Ultimate physical limits to computation." *Nature* 406 (31 Agosto): 1047-1054 (2000).

[6] H. Moravec, *Mind Children*, Harvard University Press (1989).

[7] Bostrom (1998), op. cit.

[8] Vedere i riferimenti alle note precedenti.

Pertanto, possiamo utilizzare la potenza di elaborazione richiesta per simulare il sistema nervoso centrale, come una stima del costo totale computazionale per simulare la mente umana.

Se l'ambiente è incluso nella simulazione, ciò richiederà un'ulteriore potenza di calcolo; quanta sia, dipende dalla portata e dalla granularità della simulazione. Simulare l'intero universo a livello quantico non è ovviamente fattibile, a meno che non venga radicalmente scoperta una nuova fisica. Ma per ottenere una simulazione realistica dell'esperienza umana, è necessario molto meno - basta solo tutto ciò che è necessario per garantire agli umani simulati, che interagendo con il proprio ambiente simulato mediante i consueti modi umani, non si accorgano di eventuali irregolarità. La struttura microscopica della parte interna della Terra può essere sicuramente omessa. Gli oggetti astronomici distanti possono essere sostituiti da rappresentazioni altamente compresse: la verosimiglianza necessita di estendere una banda stretta di proprietà, che potremo osservare dalla nostra navicella pianeta o sistema solare. Gli oggetti macroscopici, presenti sulla superficie della Terra nelle zone abitate, possono avere bisogno di essere continuamente simulati, ma i fenomeni microscopici potrebbero probabilmente essere riempiti *ad hoc*. Quello che si vede attraverso un microscopio elettronico, deve essere guardato senza sospetti, ma di solito non c'è modo di confermare la sua coerenza con le parti inosservate del mondo microscopico. Nascono delle eccezioni quando progettiamo deliberatamente sistemi per sfruttare i fenomeni microscopici inosservati che operano secondo principi noti per ottenere risultati che si potranno verificare in modo indipendente. Un caso paradigmatico di questo genere è il computer. La simulazione può quindi avere bisogno di includere una continua rappresentazione al computer a livello della logica individuale degli elementi. Questo non presenta alcun problema, dato che la nostra potenza di calcolo corrente, è insignificante per gli standard post-umani.

Inoltre, un simulatore post-umano potrebbe avere sufficiente potenza di calcolo per tenere una traccia dettagliata di tutte le convinzioni in tutti i cervelli umani in ogni momento. Pertanto, quando si è visto che un essere umano stava per fare un'osservazione sul mondo microscopico, si sarebbe potuta riempire la simulazione in modo sufficientemente dettagliato e nel dominio appropriato, in base alle necessità. Se si verifica un errore, il responsabile potrebbe facilmente modificare gli stati di qualsiasi cervello, che sono diventati consapevoli dell'anomalia, prima che rovini la simulazione. In alternativa, il responsabile potrebbe tornare indietro di qualche secondo e rieseguire la simulazione in modo da evitare il problema.

Sembra quindi plausibile che il principale costo computazionale nel creare delle simulazioni che sono indistinguibili dalla realtà fisica per le menti umane dentro la simulazione, sia quello di simulare cervelli organici a livello neuronale o sub-neuronale.[9] Mentre non è possibile ottenere una stima molto precisa sul costo di una simulazione realistica della storia umana, possiamo usare ~ 1033 - 1036 operazioni

[9] Come costruiamo computer sempre più veloci, il costo per simulare le nostre macchine potrebbe alla fine arrivare a dominare il costo per simulare il sistema nervoso.

come stima approssimativa.[10] Come guadagneremo più esperienza con la realtà virtuale, avremo una migliore comprensione delle esigenze di calcolo necessarie per fare apparire realistici tali mondi ai loro visitatori. Ma in ogni caso, anche se la nostra stima è sbagliata per diversi ordini di grandezza, questo non ha molta importanza riguardo a quello che stiamo dicendo. Abbiamo notato che un'approssimazione sulla potenza di calcolo di un computer a massa planetaria, è di 1042 operazioni al secondo e che presuppone solo disegni nano tecnologici già conosciuti, che probabilmente sono ben lungi dall'essere ottimali. Un singolo computer di quel genere potrebbe simulare l'intera storia mentale del genere umano (ciò che si chiama *simulazione ancestrale*) utilizzando meno di un milionesimo della sua potenza di elaborazione al secondo. Una civiltà post-umana potrà finalmente costruire un numero impressionante di quei computer. Possiamo concludere dicendo che la potenza di calcolo disponibile a una civiltà post-umana sarà sufficiente per eseguire un numero enorme di simulazioni ancestrali, persino assegnando solo una minima parte delle proprie risorse a tale scopo. Possiamo trarre questa conclusione anche lasciando un notevole margine di errore in tutte le nostre stime.

- Le civiltà post-umane avrebbero sufficiente potenza di calcolo per eseguire una quantità enorme di simulazioni ancestrali, persino utilizzando solo una piccola frazione delle loro risorse a tale scopo.

IV. IL NOCCIOLO DELL'ARGOMENTO SIMULAZIONE

L'idea di base di questa scritto può essere espressa approssimativamente come segue: Se ci fosse una sostanziale probabilità che la nostra civiltà possa arrivare alla fase post-umana ed eseguire molte simulazioni ancestrali, allora come mai non stiamo vivendo in una simulazione del genere?

Svilupperemo questa idea discutendone precisamente. Cerchiamo di introdurre le seguente voci:

f_p: La frazione di tutte le civiltà tecnologiche a livello umano che sopravvivono per raggiungere uno stadio post-umano

N: Numero medio di simulazioni ancestrali eseguite da una civiltà post-umana

H: Numero medio di individui che hanno vissuto in una civiltà prima che raggiunga una fase post-umana

[10] 100 miliardi di umani X 50 anni/umani X 30 milioni di sec/anni X [10^{14}, 10^{17}] operazioni in ogni cervello umano al secondo ≈ [10^{33}, 10^{36}] operazioni.

La frazione effettiva di tutti gli osservatori con esperienze di tipo umano che vivono in realtà simulate è quindi

$$f_{sim} = \frac{f_p\, NH}{(f_p\, NH) + H}$$

Scrivendo f_I per la frazione delle civiltà post-umane che sono interessate ad eseguire simulazioni ancestrali (o che contengono almeno alcuni individui che sono interessati a questo e dispongono di risorse sufficienti per eseguire un numero significativo di tali simulazioni), e N_I per il numero medio di simulazioni ancestrali eseguite da tali civiltà interessate, otteniamo

$$N = f_I\, N_I$$

e quindi

$$f_{sim} = \frac{f_p\, f_I\, N_I}{(f_p\, f_I\, N_I) + 1} \qquad (*)$$

A causa della immensa potenza di calcolo delle civiltà post-umane, N_I è estremamente grande, come abbiamo visto nel paragrafo precedente. Ispezionando (*) si può quindi vedere che almeno una delle seguenti tre proposizioni deve essere vera:

1) $f_p \approx 0$
2) $f_I \approx 0$
3) $f_{sim} \approx 1$

V. UN BLANDO PRINCIPIO DI INDIFFERENZA

Siamo in grado di fare un ulteriore passo avanti e concludere riguardo alla verità della condizione 3), dicendo che il credere nell'ipotesi di essere in una simulazione dovrebbe essere prossimo all'unità. Più in generale, se sapessimo che una frazione x di tutti gli osservatori con esperienze di tipo umano vivono già in simulazioni, e non abbiamo alcuna informazione che indica che le nostre esperienze particolari sono più o meno verosimili rispetto ad altre esperienze di tipo umano che

sono state implementate *in vivo* piuttosto che *in machina,* allora il nostro credere di trovarci in una simulazione deve essere x uguale a:

$$Cr(SIM \mid f_{sim} = x) = x \qquad (\#)$$

Questo passo è sancito da un principio di indifferenza molto debole. Proviamo a distinguere i due casi. Il primo caso, che è il più semplice, è dove tutte le menti in questione sono proprio come la vostra, nel senso che sono esattamente e qualitativamente identiche alla vostra: hanno esattamente le stesse informazioni e le stesse esperienze che avete voi. Il secondo caso è dove le menti sono "come" quelle di uno qualunque, nel senso che vagamente sono il tipo di mente tipica delle creature umane, ma qualitativamente sono distinte tra loro e ciascuna ha una serie distinta di esperienze. Ritengo che anche in quest'ultimo caso, in cui le menti sono qualitativamente diverse, l'argomento simulazione possa funzionare ancora, a condizione che non siano informazioni che portano alla questione: quali sono le menti simulate e quali sono quelle implementate biologicamente.

La letteratura può fornire una difesa dettagliata riguardo un principio molto più forte, la quale implica, per entrambi i casi, la stessa presa di posizione di esempi particolari insignificanti.[11] Lo spazio a disposizione non consente una ricapitolazione di detta difesa in questa sede, ma siamo in grado di mettere in evidenza una delle intuizioni di fondo portando la nostra attenzione su una situazione analoga in un genere più familiare. Supponiamo che x% della popolazione ha una certa sequenza genetica S all'interno del proprio DNA comunemente conosciuto come "DNA spazzatura". Inoltre, supponiamo che non vi siano manifestazioni di S (abbreviazione per tutto ciò che potrebbe virare nel gene sotto analisi) e che non vi siano correlazioni note tra avere un S e qualsiasi altra caratteristica osservabile. Quindi chiaramente, a meno che non avete sequenziato il vostro DNA, è razionale assegnare un credito di x% all'ipotesi che avete il gene S. E ciò è così a prescindere completamente dal fatto che le persone che hanno S abbiano, sia la mente che le esperienze, qualitativamente diverse dalle persone che non hanno S. (Sono differenti semplicemente perché tutti gli esseri umani hanno diverse esperienze gli uni dagli altri, non a causa di un qualsiasi link conosciuto fra S e il tipo di esperienze che uno ha.)

Lo stesso ragionamento vale se S non è la proprietà richiesta per avere una certa sequenza genetica, ma invece la proprietà di essere in una realtà simulata, ipotizzando solo di non avere le informazioni che ci consentono di prevedere eventuali differenze tra le esperienze delle menti simulate e quelle delle menti originali biologiche.

Va sottolineato che il blando principio di indifferenza espresso da (#) prescrive solo l'indifferenza tra le ipotesi riguardo a *che tipo di osservatori siete* e *quando non disponete di alcuna informazione su quale tipo di osservatori siete.*

[11] Per esempio N. Bostrom in "The Doomsday Argument, Adam & Eve, UN++, and Quantum Joe." *Sintesi* 127 (3): 359-387 (2001); e maggiormente nel mio libro *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, Rouledge, New York, 2002.

In generale, non indica indifferenza tra le ipotesi in cui vi mancano specifiche informazioni riguardo a quale sia quella vera. Essendo in contrasto con il principio di indifferenza di Laplace e altri più ambiziosi, è quindi immune al paradosso di Bertrand e ad altre situazioni difficili che tendono ad affliggere i principi di indifferenza di portata illimitata.

I lettori che hanno familiarità con l'argomento Fine del Mondo (DOOM)[12] potrebbero essere preoccupati che il blando principio di indifferenza invocato qui sia la stessa ipotesi responsabile di far decollare l'argomento Fine del Mondo, e che la contro intuitività di alcune tra le implicazioni di quest'ultima, incrimina o fa dubitare sulla validità della prima. Ma non è così. L'argomento Fine del Mondo poggia su una premessa *molto più* forte e più controversa, vale a dire che dovrebbe essere visto come se uno fosse un campione casuale dal set di tutte le persone che hanno vissuto da sempre (passate, presenti e future), *sebbene sappiamo di vivere all'inizio del XXI secolo*, piuttosto che in un certo punto nel lontano passato o nel futuro. Il principio blando di indifferenza, al contrario, si applica solo ai casi in cui non si hanno informazioni su quale gruppo di persone si appartiene.

Anche se le quote delle scommesse forniscono un orientamento razionale, merita una riflessione il fatto che se ciascuno dovesse piazzare una scommessa sull'essere in una simulazione o no, userebbe il principio blando di indifferenza e di conseguenza punterebbe i propri soldi sull'essere in una simulazione, in quanto sa che è dove puntano quasi tutti, e quindi quasi tutti vincerebbero la scommessa. Se invece scommetteranno di non essere in una simulazione, quasi tutti perderanno. Sembra che sia meglio dar retta al principio blando di indifferenza.

Inoltre, possiamo considerare una sequenza di situazioni possibili nelle quali aumenti la frazione di tutte le persone che vivono nelle simulazioni: 98%, 99%, 99,9%, 99,9999%, e così via. Come ci avviciniamo al caso limite in cui tutti sono in una simulazione (da cui uno può *dedurre* di essere pure lui stesso in una simulazione), è plausibile richiedere che il credito abbinato all'essere in una simulazione, si avvicini gradualmente in modo da corrispondere al caso limite di assoluta certezza.

VI. INTERPRETAZIONE

La possibilità rappresentata dalla proposizione (1) è abbastanza semplice. Se (1) è vera, allora il genere umano quasi certamente non riuscirà a raggiungere un livello post-umano; praticamente nessuna specie al nostro livello di sviluppo diventerà post-umana, ed è difficile avere una qualsiasi giustificazione che ci permetta di pensare che la nostra specie sarà particolarmente privilegiata o protetta da disastri futuri. Perciò, la condizione posta da (1) è che dobbiamo dare un alto credito alla Fine del Mondo

12 Vedere J. Leslie "Is the End of the World Night?" *Philosophical Quarterly* 40, 158: 65-72 (1990).

(DOOM), ossia ipotizzare che l'umanità si estinguerà prima di raggiungere un livello post-umano:

$$Cr(\text{DOOM} \mid f_p \approx 0) \approx 1$$

Si possono immaginare delle situazioni ipotetiche in cui siamo in possesso di prove incredibili, tipo ottenere la conoscenza di f_p. Per esempio, se scopriamo di essere in procinto di venire colpiti da un meteorite gigante, questo potrebbe suggerire che siamo stati eccezionalmente sfortunati. Potremmo quindi assegnare un credito a DOOM più grande della nostra aspettativa sulla frazione di civiltà a livello umano che non riesce a raggiungere la post-umanità. Tuttavia, nel caso attuale ci sembra che manchino le prove per pensare di essere speciali in questo senso, sia nel bene che nel male.

La proposizione (1) di per sé non implica che ci siano le probabilità di estinguersi presto, ma solamente che è improbabile raggiungere una fase post-umana. Questa possibilità è compatibile col fatto di restare al nostro attuale livello di sviluppo tecnologico (o qualcosa in più), per molto tempo prima di estinguerci. Un altro modo per cui (1) può essere vera, è se è probabile che la civiltà tecnologica crollerà. Le società umane primitive potrebbero quindi rimanere sulla Terra per un tempo indeterminato.

Ci sono molti modi in cui l'umanità potrebbe estinguersi prima di raggiungere la post-umanità. Forse, l'interpretazione più naturale di (1) è che rischiamo di estinguerci a causa dello sviluppo di qualche potente, ma pericolosa, tecnologia.[13] Un candidato è la nanotecnologia molecolare, che nella sua fase matura consentirebbe la costruzione di nanobot auto replicanti capaci di nutrirsi di sporcizia e di materia organica - una specie di batteri meccanici. Tali nanobot progettati per fini malevoli, potrebbero causare l'estinzione di tutta la vita sul nostro pianeta.[14]

La seconda alternativa, in conclusione dell'argomento simulazione, è che la frazione delle civiltà post-umane che sono interessate a eseguire simulazioni ancestrali, sia trascurabile. Affinché (2) possa essere vera, ci deve essere una forte *convergenza* nelle varie direzioni delle civiltà avanzate. Se il numero delle simulazioni ancestrali create dalle civiltà interessate è molto grande, la rarità di tali civiltà deve essere corrispondentemente estrema. Praticamente, nessuna civiltà post-umana decide di utilizzare le proprie risorse per eseguire un certo numero di simulazioni ancestrali. Inoltre,

---

[13] Vedere il mio scritto "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards." *Journal of Evolution and Technology*, vol. 9 (2001) per un'indagine e un'analisi del presente e delle minacce future previste per la sopravvivenza umana.

[14] Vedere Drexler (1985), cit. op. e R. A. Freitas Jr., "Some Limits to Global Ecophagy by Biovorous Nanoreplicators, with Public Policy Recommendations." *Zyvex preprint* Aprile (2000), http://www.foresight.org/NanoRev/Ecophagy.html

in quasi tutte le civiltà post-umane c'è carenza di individui che hanno le risorse e l'interesse per eseguire sufficienti simulazioni ancestrali; oppure hanno applicato in modo affidabile le leggi in modo da evitare che tali individui agiscano in base ai loro desideri.

Quale forza potrebbe causare un tale convergenza? Si può ipotizzare che tutte le civiltà avanzate si siano sviluppate lungo una traiettoria che conduce al riconoscimento di un divieto etico contro l'esecuzione di simulazioni ancestrali a causa della sofferenza che sarebbe inflitta agli abitanti della simulazione. Tuttavia, dal nostro punto di vista attuale, non è chiaro che sia immorale la creazione di una razza umana. Al contrario, tendiamo a vedere l'esistenza della nostra razza come un costituente di grande valore etico. Inoltre, la convergenza sulla visione etica della immoralità nell'eseguire simulazioni ancestrali non è sufficiente: deve essere combinata alla convergenza della struttura sociale a livello di civiltà che consenta di vietare le attività considerate immorali.

Un altro possibile punto di convergenza è che quasi tutti gli individui post-umani in quasi tutte le civiltà post-umane, si sviluppano in una direzione in cui perdono i loro desideri nell'eseguire tali simulazioni. Ciò richiederebbe delle modifiche significative alle motivazioni che hanno spinto i loro predecessori umani, perché ci sono certamente molti esseri umani che vorrebbero eseguire tali simulazioni se potessero permettersi di farlo. Ma forse molti dei nostri desideri umani saranno considerati stupidi da chiunque sia diventato post-umano. Può essere che il valore scientifico delle simulazioni ancestrali sia trascurabile per una civiltà post-umana (il che non è tanto improbabile data la sua insondabile superiorità intellettuale), e forse i post-umani considerano le attività ricreative semplicemente come un modo molto inefficiente per ottenere piacere - che si può ottenere molto più economicamente da una diretta stimolazione dei centri edonici del cervello. Una conclusione che ne consegue da (2) è che le società post-umane saranno molto diverse dalle società umane: non conterranno agenti indipendenti relativamente ricchi che hanno l'intera gamma dei desideri umani e sono liberi di soddisfarli.

La possibilità espressa dall'alternativa (3) è quella più concettualmente intrigante. Se già stiamo vivendo in una simulazione, il cosmo che stiamo osservando è quindi solo un piccolo pezzo della totalità dell'esistenza fisica. La fisica dell'universo in cui si trova il computer che esegue la simulazione può e non può assomigliare alla fisica del mondo che osserviamo. Sebbene il mondo che vediamo è in un certo senso "reale", non è situato al livello basico della realtà.

Potrebbe essere possibile, alle civiltà simulate, di diventare post-umane. Quindi potranno eseguire le loro simulazioni ancestrali su potenti computer che costruiranno nel loro universo simulato. Tali computer saranno delle "macchine virtuali", un concetto familiare in informatica. (Il web applet Java script, per esempio, funziona su una macchina virtuale - un computer simulato situato all'interno del vostro desktop) Le macchine virtuali possono essere ammassate: è possibile simulare una macchina che simula un'altra macchina e via dicendo, arbitrariamente in questo modo e in molti passi di iterazione.

Se non creeremo le nostre simulazioni ancestrali, questo costituirebbe una forte prova contro (1) e (2), e perciò dovremmo quindi concludere affermando di vivere in una simulazione. Inoltre, dovremmo sospettare che i post-umani che stanno eseguendo la loro simulazione, siano essi stessi degli esseri simulati; anche i loro creatori, a loro volta, possono essere degli esseri simulati.

La realtà può quindi contenere molti livelli. Anche se è necessario che la gerarchia, a un certo punto, tocchi il fondo - lo status metafisico di questa affermazione è alquanto oscuro - ci può essere spazio per un gran numero di livelli di realtà e il numero potrebbe aumentare nel corso del tempo. (Una considerazione che va contro all'ipotesi dei multi livelli, è che il costo computazionale per i simulatori del livello base, sarebbe molto grande. Simulare anche solo una civiltà post-umana potrebbe essere proibitivo. Se è così, allora dovremmo aspettarci che la nostra simulazione termini quando stiamo per diventare post-umani.)

Sebbene tutti gli elementi di un tale sistema possano essere naturalistici e persino fisici, è possibile trarre alcune vaghe analogie con le concezioni religiose del mondo. In un certo senso i post-umani che eseguono una simulazione, rispetto alle persone che abitano nella simulazione, sono come degli dei: i post-umani hanno creato il mondo che vediamo; sono di intelligenza superiore; sono "onnipotenti" nel senso che possono interferire con il funzionamento del mondo, anche utilizzando dei modi che violano le sue leggi fisiche; e sono "onniscienti" nel senso che possono monitorare tutto ciò che accade. Tuttavia, tutti i semidei, ad eccezione di quelli al livello primario della realtà, sono soggetti a punizioni da parte delle più potenti divinità che vivono nei livelli inferiori.

Ulteriori elucubrazioni su questi temi potrebbero raggiungere l'apice in una teogonia naturalistica che aiuterebbe a studiare la struttura di questa gerarchia ei vincoli imposti ai suoi abitanti, tramite la possibilità che le azioni svolte al proprio livello possano influire sul trattamento ricevuto dagli abitanti dei livelli più profondi. Ad esempio, se nessuno può essere sicuro di essere al livello del sottosuolo, è ovvio che ognuno potrebbe prendere in considerazione la possibilità che le proprie azioni vengano premiate o punite dai propri simulatori, magari in base a criteri morali. L'aldilà sarebbe una possibilità reale. A causa di questa incertezza fondamentale, anche la civiltà del sottosuolo può avere un motivo per comportarsi in modo etico. Il fatto di avere un motivo per tenere un comportamento morale, ovviamente aggiungerebbe un motivo in più a tutti gli altri per comportarsi moralmente e così via, formando un cerchio veramente virtuoso. Si potrebbe ottenere una sorta di imperativo etico universale, la cui obbedienza sarebbe nell'interesse di tutti, come se giungesse "dal nulla". Oltre alle simulazioni ancestrali, si possono anche prendere in considerazione la possibilità di eseguire simulazioni più selettive che comprendono solo un piccolo gruppo di umani o un singolo individuo.

Il resto dell'umanità sarebbe quindi a livello di zombie o "uomini ombra", cioè esseri umani simulati creati solo per vivere a un livello sufficiente per non accorgersi di qualcosa di sospetto. Non è chiaro quanto più economico sia simulare uomini ombra rispetto a uomini veri. Non è neppure ovvio che sia possibile per un soggetto di comportarsi indiscutibilmente da essere umano reale, mancando tuttavia di esperienze coscienti. Anche se ci sono tali simulazioni selettive, non dovete pensare di trovarvi in una di loro a meno che non pensiate che siano molto più numerose delle simulazioni complete. Ci dovrebbero essere circa 100 miliardi di volte in più di "simulazioni del me" (simulazioni della vita di una singola mente) di quante siano le simulazioni ancestrali, per far sì che la maggior parte delle persone simulate sia in una simulazione del me.

C'è anche la possibilità che dei simulatori limitino certe parti della vita mentale degli esseri simulati, dando loro false memorie tipo di esperienze che avrebbero tipicamente avuto durante l'intervallo omesso. Se è così, si può considerare la seguente soluzione (inverosimile) al problema del male: che non c'è sofferenza nel mondo e tutti i ricordi di sofferenza sono illusioni. Naturalmente, questa ipotesi può essere seriamente concepita solo in quei momenti in cui non si sta veramente soffrendo.

Supponendo di stare vivendo in una simulazione, quali sono le implicazioni per noi esseri umani? Quanto sopra sottolinea, che malgrado tutto, le implicazioni non sono tutte radicali. La nostra migliore guida su come i nostri creatori post-umani abbiano scelto di impostare il nostro mondo, è lo studio empirico e standard dell'universo che vediamo. Le revisioni della maggior parte dei nostri network di credi e convinzioni, sarebbero piuttosto insignificanti e sfuggenti, in proporzione alla nostra mancanza di fiducia nella nostra capacità di comprendere le modalità post-umane. Pertanto, una volta compresa correttamente, la verità di (3) dovrebbe tendere a farci "impazzire" o impedirci di svolgere la nostra attività e fare progetti e previsioni per il domani. L'importanza empirica dominante di (3), al momento attuale sembra risiedere nel suo proprio ruolo stabilito nella conclusione tripartita sopra citata.[15] Possiamo sperare che (3) sia vera in quanto ciò diminuirebbe la probabilità di (1), anche se i vincoli di calcolo rendono probabile che i simulatori potrebbero terminare la simulazione prima che raggiunga un livello post-umano, per cui la nostra migliore speranza sarebbe che (2) sia vera.

Se impariamo di più sui vincoli delle risorse e sulle motivazioni per una post-umanità, forse come risultato di sviluppi volti a diventare noi stessi post-umani, allora l'ipotesi che siamo esseri simulati arriverà ad avere un insieme molto più ricco di implicazioni empiriche.

---

15 Per alcune riflessioni di un altro autore sulle conseguenze di (3), che sono state accese da una precedente versione di questo documento circolata privatamente, vedere R.Hanson, “How to Live in a Simulation”, *Journal of Evolution and Technology*, vol.7 (2001).

VII. CONCLUSIONI

Una civiltà "post-umana" pienamente sviluppata tecnologicamente avrebbe enormi potenze di calcolo. Sulla base di questo fatto empirico, l'argomento simulazione mostra che almeno una delle seguenti affermazioni è vera: (1) La frazione delle civiltà a livello umano che raggiungono uno stadio post-umano è molto vicina allo zero; (2) La frazione delle civiltà post-umane che sono interessate ad eseguire simulazioni ancestrali è molto vicina allo zero; (3) La frazione di tutti gli individui con il nostro tipo di esperienze che vivono in una simulazione è molto vicina a uno.

Se (1) è vera, allora quasi certamente ci estingueremo prima di raggiungere la post-umanità. Se (2) è vera, allora ci deve essere una forte convergenza tra le direzioni delle civiltà avanzate in modo che praticamente nessuna contenga individui relativamente ricchi che desiderano gestire tali simulazioni e siano liberi di farlo. Se (3) è vera, allora quasi certamente stiamo vivendo in una simulazione. Nella foresta oscura della nostra ignoranza attuale, sembra ragionevole ripartire più o meno equamente le nostre convinzioni tra (1), (2) e (3).

A meno che non stiamo vivendo in una simulazione, i nostri discendenti quasi certamente non eseguiranno mai una simulazione ancestrale.

Riconoscimenti

Sono grato alle molte persone per i loro commenti, e in particolare ad Amara Angelica, Robert Bradbury, Milan Cirkovic, Robin Hanson, Hal Finney, Robert A. Freitas Jr., John Leslie, Mitch Porter, Keith DeRose, Mike Treder, Mark Walker, Eliezer Yudkowsky, e ai molti altri commentatori anonimi.