

Inicio: [Ensayos sobre la reducción del sufrimiento](#)

Riesgos de sufrimiento astronómico en el futuro

por Brian Tomasik

Escrito originalmente: dic. 2011; cambios importantes: dic. 2012, mar. 2013, oct. 2013;
última actualización: 8 ene. 2014. Traducción de Sergio Tarrero

1. Resumen

No está nada claro que los valores humanos vayan a mantenerse y a dar forma a una onda expansiva de colonización espacial originada en la Tierra, pero incluso si lo hacen, parece más probable que la colonización del espacio aumentará el sufrimiento total en vez de disminuirlo. Dicho esto, como hay otras personas con un gran interés en la supervivencia y expansión de la humanidad por el cosmos, soy de la opinión de que quienes buscamos reducir el sufrimiento deberíamos dejar que los demás intenten cumplir sus sueños acerca de la expansión espacial a cambio de medidas de seguridad más estrictas contra el sufrimiento futuro.

▪

2. Índice

- [1. Resumen](#)
- [2. Índice](#)
- [3. Epígrafes](#)
- [4. Es posible que los valores humanos no controlen el futuro](#)
- [5. Algunos escenarios de sufrimiento futuro](#)
 - [5.1. Proliferación de animales salvajes](#)
 - [5.2. Simulaciones sintientes](#)
 - [5.3. Subrutinas que sufren](#)

- [5.4. Cisnes negros \(eventos inesperados de grandes consecuencias\)](#)
- [6. Incluso un futuro controlado por humanos es altamente probable que incremente el sufrimiento](#)
 - [6.1. Proliferación de animales salvajes](#)
 - [6.2. Simulaciones sintientes](#)
 - [6.3. Subrutinas que sufren](#)
 - [6.4. Cisnes negros](#)
- [7. ¿Y qué pasa con los productores de clips?](#)
- [8. Por qué deberíamos continuar teniendo una actitud de cooperación](#)
- [9. Posibles aspectos positivos de un futuro inteligente](#)
 - [9.1. Cisnes negros que no sean armas de doble filo](#)
 - [9.2. Valorando la reflexión](#)
- [10. Agradecimientos](#)
- [11. Notas a pie de página](#)

3. Epígrafes

“Si llevamos la antorcha verde de una estrella a otra, y encendemos en torno a cada una conflagración de vitalidad, podemos desencadenar una metamorfosis Universal. [...] Por nuestra actividad [...] la escoria se convertirá en tierra, hierba brotará, flores florecerán y bosques emergerán en lugares que una vez fueron estériles. ^[1] [...] Si negamos nuestro increíble desafío; damos la espalda al universo vivo, y abandonamos nuestro destino cósmico, vamos a cometer un crimen de magnitud indescriptible.” – Marshall T. Savage, *El Proyecto Milenio: La colonización de la galaxia en ocho pasos fáciles*, 1994 (de "[Citas del Espacio](#)")

“Oremos para que la raza humana nunca se escape de la Tierra para difundir su iniquidad a otra parte.” – [CS Lewis](#)

“Si no puedes con ellos, únete a ellos.” – [proverbio](#)

4. Es posible que los valores humanos no controlen el futuro

El ensayo de Nick Bostrom "[El futuro de la evolución humana](#)" describe un escenario en el que los valores humanos de la diversión, el ocio y las relaciones podrían ser reemplazados por agentes hiperoptimizados que competirían mejor en la carrera darwiniana para controlar nuestro cono de luz futuro. La única forma de poder evitar este escenario competitivo, sugiere Bostrom, sería a través de un "[singleton](#)", un agente unificado o una estructura de gobierno que pudiera controlar la evolución. Por supuesto, puede que un *singleton* no mantenga los valores humanos. Muchos agentes ingenuos de IA (inteligencia artificial) que los humanos podrían construir tendrían la capacidad de [optimizar una función objetiva que los seres humanos encuentran sin sentido](#). O incluso si los humanos mantenemos las manos al volante, estamos lejos de poder garantizar que podamos conservar nuestros objetivos de manera modificaciones futuras más significativas.

Estos factores sugieren que incluso si el progreso tecnológico humano continúa, la probabilidad de que los valores humanos se mantengan en el futuro puede no ser muy alta. Para que los valores humanos se hagan realidad parece que es necesario un *singleton* que no sea un optimizador ciego, que pueda preservar de manera estable los valores, y que sea conformado por diseñadores que se preocupen por los valores humanos, en vez de por beneficios egoístas u otras cosas. Es importante recordar esto cuando imaginamos lo que los seres humanos futuros podrían ser capaces de conseguir con su tecnología.

Algunas personas creen que si llegan a existir superinteligencias suficientemente avanzadas estas descubrirán la verdad moral, y, en consecuencia, harán necesariamente lo correcto. Por lo tanto, se afirma, si la humanidad sobrevive y se hace más inteligente al final alguien hará aquello que es correcto. Este punto de vista es problemático por dos motivos

. En primer lugar, la aplicación de la navaja de Occam lleva a cuestionar la existencia de una verdad moral (sea lo que sea que significa dicha existencia). En segundo lugar, incluso si esa verdad moral existiese, ¿por qué debería importarle a una superinteligencia? Hay un montón de personas brillantes en la Tierra hoy en día que comen carne. Saben perfectamente el sufrimiento que causa, pero sus sistemas motivacionales no están lo

suficientemente afectados por el daño que están causando a los animales que se encuentran en las granjas. Lo mismo puede ser cierto para superinteligencias. De hecho, podemos pensar en inteligencias arbitrarias en el conjunto de las inteligencias posibles que [no tendrían por qué tener](#) el más mínimo indicio de empatía por el sufrimiento que experimentan los seres sintientes.

5. Algunos escenarios de sufrimiento futuro

Incluso si los humanos conservar el control sobre el futuro de la vida basada en la Tierra hay muchas maneras mediante las cuáles la colonización del espacio multiplicaría el sufrimiento. Estas son algunas de ellas.

5.1. Propagación de animales salvajes

Los humanos podrían colonizar otros planetas, extendiendo la [vida animal](#), llena de sufrimiento, a través de [la terraformación](#). Algunos seres humanos podrían utilizar sus recursos para [sembrar la vida por toda la galaxia](#), lo que, tristemente, algunos consideran un [imperativo moral](#).



5.2. Simulaciones sintientes

Dada una capacidad computacional [astronómica](#), los post-humanos podrían ejecutar simulaciones de sus antepasados (incluidas las cámaras de tortura, los campos de exterminio, y las enfermedades psicológicas que sufren miles de millones de personas). Estas simulaciones podrían incluir también muchas copias de animales salvajes, la mayoría de los cuales mueren dolorosamente poco después de nacer. Lo que es más, los

científicos podrían ejecutar un número aún mayor de simulaciones de “organismos que podrían haber existido”, explorando el espacio de mentes posibles. Podrían simular sextillones de “aprendices por refuerzo” que fuesen suficientemente autoconscientes como para sentir lo que consideramos dolor consciente.

5.3. Subrutinas de sufrimiento

[Podría darse el caso](#) de que ciertos algoritmos (digamos, por ejemplo, unos simples [agentes de refuerzo](#)) resultasen ser muy útiles para realizar cálculos complejos de aprendizaje-máquina, que tuviesen que ser ejecutados por una inteligencia artificial (IA) avanzada y a una escala enorme. Estas subrutinas podrían acabar siendo lo suficientemente similares a los programas de dolor en nuestros cerebros para finalmente acabar sufriendo de verdad. Pero como la búsqueda de beneficios y el poder se impondrían por encima de la compasión, estos subprogramas podrían acabar siendo ampliamente usados en los cerebros matrioska de la IA [N. del T.: un cerebro matrioska es una megaestructura hipotética de inmensa capacidad computacional].

5.4. Cisnes negros (eventos impredecibles de grandes consecuencias)

La gama de escenarios que podemos imaginar es limitada, y pueden surgir muchas más posibilidades que no se nos han ocurrido, o que tal vez ni siquiera podamos comprender. [N. del T.: la expresión “cisne negro” se usa para nombrar un evento atípico que a pesar de tener un impacto muy notable no podría haber sido predicho].

6. Incluso un futuro humano controlado es probable que aumente el sufrimiento

Si tuviese que estimarlo ahora mismo, yo daría un ~65% a la probabilidad de que si los seres humanos decidiesen colonizar el espacio, esto causaría por razones intrínsecas (esto es, si dejamos de lado las consideraciones acerca de las posibilidades de llegar a acuerdos que [se explican más adelante](#)) más sufrimiento del que reduciría. Pensemos en las maneras en las que la colonización del espacio podría *reducir* el sufrimiento de alguna forma plausible. Para la mayor parte de los mecanismos que tendrían tal efecto, parece que

existen también contra-mecanismos que *aumentarían* el sufrimiento por lo menos al mismo nivel. Las siguientes secciones son paralelas a las anteriores.

6.1. *Propagación de animales salvajes*

David Pearce [acuñó la frase](#) "misiones de rescate cósmico" para referirse a la posibilidad de enviar sondas a otros planetas para aliviar el sufrimiento salvaje extraterrestre que estos pudiesen contener. Esta es una buena idea, pero es problemática por algunas razones.

- No hemos encontrado ningún extraterrestre, por el momento, así que no es obvio que haya un gran número de ellos esperando ser rescatados de su agonía darwiniana.
- No tendríamos por qué encontrar necesariamente entre los extraterrestres el tipo específico de sufrimiento consciente que nos es tan conocido en la vida animal terrestre. Los extraterrestres probablemente sean bacterias, plantas, etc., e incluso si son inteligentes, podrían ser inteligentes a la manera en la que son inteligentes los robots, sin emociones como las que tanto nos importan. (Sin embargo, si son muy sofisticados, sería [relativamente poco probable](#) que no los considerásemos conscientes).
- El viaje espacial es lento y difícil.
- No está claro que la humanidad fuese a apoyar este tipo de misiones. Los ecologistas nos pedirían que dejásemos los hábitats extraterrestres en paz. Habría otros que no querrían gastar recursos en misiones de rescate, a no ser que planeasen explotar los recursos naturales de esos planetas.

Comparemos esto con las posibilidades de *propagación* del sufrimiento de animales salvajes:

- Los seres humanos pueden propagar la vida a muchos planetas (por ejemplo, a través de la terraformación de Marte, o a otros planetas similares a la Tierra a través de la panspermia dirigida). El número de planetas que puedan albergar vida podría

ser considerablemente mayor que el número de los que ya la albergan. (Véase la discusión de f_i en la [ecuación de Drake](#)).

- A diferencia de la vida extraterrestre, ya sabemos que hay vida terrestre sintiente.
- Propagar la vida biológica es lento y difícil, pero lanzar pequeñas cápsulas con la capacidad de producir vida es más fácil que enviar sondas para atender el Imperativo Hedonista, o [sondas Berserker](#).
- Afortunadamente, es posible que los seres humanos no apoyen mucho la propagación de la vida, aunque [algunos lo hacen](#). Respecto a la terraformación, a corto plazo tenemos presiones existenciales o de supervivencia obvias que nos instan a hacerlo, pero probablemente la panspermia dirigida sea un problema mayor a largo plazo.

Aunque arriba apunté que no es seguro que los extraterrestres sean sintientes, realmente creo que es moderadamente probable que la conciencia sea bastante convergente entre civilizaciones inteligentes.

Esto se basa en (a)

la [apreciación](#) de una conciencia convergente entre animales terrestres y (b) el principio



general de que la conciencia parece ser útil para [la planificación, la manipulación de imágenes, el automodelado, etc.](#) Por otro lado, esto tal vez sea reflejo de la precariedad de mi imaginación humana a la hora de concebir formas en las que se puede ser inteligente sin poseer conciencia.

Dicho esto, incluso si los extraterrestres son conscientes, pueden no tener emociones que nos importen. Las emociones parecen ser modos de funcionamiento que guían a un organismo a hacer ciertas cosas en determinados momentos (comer, dormir, tener relaciones sexuales, correr, gritar, etc.). También están relacionadas con el aprendizaje por

refuerzo. No parece que ninguna de estas emociones sea necesaria en una IA que no sea un agente del tipo de agentes que son los animales. Por ejemplo, imaginemos una IA que calcule los valores esperados para las acciones sin "sentir" los premios y castigos directamente. Los cambios y actualizaciones de sus planes podrían ocurrir a través de [cálculos basados en modelos, en vez de a través del aprendizaje por refuerzo estándar](#). Por supuesto, puede ser que lo que me lleva a decir que estas actualizaciones modelo más abstractas no son también una forma de recompensa/castigo no sea más que una estrechez de miras por mi parte. Pero no está claro que una IA fuese a tener el mismo tipo de reacciones hacia ella que los seres humanos tienen hacia sus actualizaciones de recompensa/castigo. Nuestra valoración moral de este caso puede depender de dónde nos situemos en el debate entre [el utilitarismo hedonista y el de las preferencias](#).

6.2. Simulaciones sintientes

Puede ser que el sufrimiento biológico sea una gota en el océano comparado con el sufrimiento digital, en cuyo caso las simulaciones de humanos podrían ser aún más preocupantes que la propagación humana de los animales salvajes.

Por supuesto, tal vez haya extraterrestres ejecutando simulaciones de la naturaleza por motivos científicos o por diversión, o simulaciones de mentes en general para estudiar la biología, la psicología y la sociología. Si nos encontrásemos con estos extraterrestres, tal vez podríamos negociar con ellos para asegurarnos de que no causan sufrimiento innecesario a sus simulaciones.

Creo que es probable que los humanos seamos más empáticos que la civilización consciente media, porque

1. parece que somos mucho más empáticos que el animal terrestre medio, probablemente en parte debido a impulsos parentales y en parte debido al comercio, aunque probablemente algunos de estos factores serían necesariamente ciertos para cualquier civilización tecnológicamente avanzada.

2. el sesgo de selección implica que estaremos más de acuerdo con la moral de nuestra sociedad que con la de otra sociedad cualquiera, porque estos son los valores con los que se nos educó y hacia los que nuestra biología nos empuja.

Basándonos en estas consideraciones, parece plausible que podría haber espacio para lograr mejoras mediante acuerdos con extraterrestres. Efectivamente, deberíamos, en general, esperar que sea posible que cualquier par de civilizaciones o facciones consigan ganancias del comercio, si tienen una utilidad marginal decreciente con respecto a la cantidad de control ejercida. Además, puede haber mejoras de Pareto baratas que se den sólo a partir de una mayor inteligencia y una mejor comprensión de consideraciones importantes, como la manera de negociar entre valores en competencia.

Dicho esto, hay algunos riesgos. Los mismos post-humanos podrían crear simulaciones que sufren y, lo que es peor, las simulaciones que los post-humanos ejecutasen sería más probable que fuesen sintientes que las ejecutadas por unos extraterrestres cualesquiera. Esto sería debido a que, de entre todas las mentes posibles, los post-humanos tendrían una tendencia a simular aquellas más cercanas a las suyas. Podrían ejecutar simulaciones de antepasados por diversión, simulaciones de la naturaleza para la apreciación estética, simulaciones de laboratorio para hacer experimentos científicos, simulaciones de mascotas para disfrutar de mascotas. Podría haber sádicos que ejecutasen simulaciones que sufriesen tormentos. En un hipotético mundo maximizador de clips, los sádicos podrían ejecutar simulaciones de clips siendo destruidos, pero eso no nos concierne. Por último, ni siquiera sabemos si hay extraterrestres accesibles para negociar sobre la reducción del sufrimiento. Sin embargo, sí que sabemos que los post-humanos probablemente ejecutarían simulaciones si colonizasen el espacio.

6.3. Subrutinas que sufren

Cualquier IA, ya sea inspirada en los seres humanos o no, podría dar como resultado la aparición de subrutinas que sufrirían. También podrían ser ejecutadas por extraterrestres, y tal vez los seres humanos podrían solicitar a los alienígenas que las diseñasen de maneras más éticas, aunque todo esto parece especulativo.

6.4. *Cisnes negros*

Parece plausible que algo totalmente inesperado pudiese tener una importancia que prevaleciese sobre la del sufrimiento en el futuro. Esto podría consistir en un nuevo descubrimiento de la física, la neurociencia, o incluso la filosofía en general. Algunos argumentan que, como sabemos tan poco en este momento, es mejor que los seres humanos continúen existiendo, por el "valor de la opción" que ello tiene. Esto es: si más adelante los seres humanos se dan cuenta de que la propagación por el espacio es mala, pueden pararla, pero si se dan cuenta de que sí deberían propagarse, pueden proceder a reducir el sufrimiento de alguna forma novedosa que no hemos anticipado.

Por supuesto, el problema con el argumento del "valor de la opción", es que asume que los seres humanos futuros harán lo correcto, cuando en realidad podemos afirmar, en base a ejemplos especulativos que podemos imaginar ahora, que parece que los humanos del futuro probablemente actuarían de forma incorrecta gran parte del tiempo. Por ejemplo, ante un nuevo descubrimiento de cantidades enormes de potencia computacional en alguna parte, la mayoría de los humanos la utilizarían para poner en marcha muchísimas más mentes. Una fracción no trivial de estas mentes podría sufrir terriblemente. En general, la mayoría de las fuentes inmensas de energía son armas de doble filo, que pueden crear más felicidad y más sufrimiento, y el impulso humano típico de promover la vida y la conciencia, en vez de eliminarlos, sugiere que quienes sostienen una posición utilitarista negativa o utilitarista con tendencia negativa están en el lado perdedor.

Aún así, el esperar y aprender más sobre esto es algo que cumple el criterio de eficiencia de Kaldor-Hicks, y tal vez haya maneras en las que puede ser también eficiente de Pareto mediante la otorgación de concesiones adicionales a los reductores de sufrimiento a modo de compensación.

7. **¿Qué pasa con los clips?**

Anteriormente he estado asumiendo, en general, una civilización orientada a la humanidad, con valores que reconocemos. Pero ¿y si, como parece ligeramente probable,

la Tierra es arrasada por un maximizador de clips, es decir, por una automatización sin restricciones ni proceso de optimización? ¿No sería esto una buena cosa, ya que eliminaría el sufrimiento salvaje, mientras los clips se propagan por toda la galaxia, sin causar ningún sufrimiento adicional?

Tal vez, pero si el maximizador de clips muestra tener una inteligencia general real, entonces no se dedicará solo a embaldosar el sistema solar con clips. Querrá hacer ciencia, aprender sobre otras mentes a través de simulaciones, posiblemente ejecutar subrutinas que sufren, y así sucesivamente. Los maximizadores de clips implicarían escenarios de cisne negro similares a una IA inspirada en el ser humano. Los maximizadores de clips probablemente serían menos intrínsecamente humanos que una "[IA amigable](#)", por lo que algunos podrían causar más sufrimiento que una IA amigable, aunque otros podrían causar menos, especialmente los "mimizadores" de clips, por ejemplo, los minimizadores de cáncer, o de muerte.

8. ¿Por qué debemos seguir siendo cooperativos?

Si el desarrollo tecnológico y la colonización del espacio parecen a punto de causar cantidades astronómicas de sufrimiento, ¿no deberíamos hacer todo lo posible para detenerlos? Bueno, vale la pena tener una discusión sobre el grado en que, como sociedad, queremos estos resultados, pero mi conjetura es que habrá alguien los continuará intentando conseguir, y esto sería difícil de frenar sin recurrir a medidas extremas. Eventualmente, quienes continúen desarrollando las tecnologías tendrán la mayor parte del poder en el mundo. Estas personas tendrán fuertes deseos de desarrollar la IA y colonizar el espacio, aunque sólo sea por el efecto de selección.

Podría ser que la resistencia no fuese completamente inútil. Hay una pequeña probabilidad de que los reductores de sufrimiento influyan en la sociedad de tal manera que se evite la colonización del espacio. Pero [sería mejor](#) para los reductores de sufrimiento llegar a un compromiso con los tecnólogos, en vez de luchar contra ellos, en esta línea: "os permitiremos adentraros en el cosmos si le dais más peso a nuestras preocupaciones sobre el sufrimiento futuro". En vez de ofrecer una pequeñísima

probabilidad de victoria completa para los reductores de sufrimiento, este enfoque de cooperación ofrece una mayor probabilidad de conseguir una fracción apreciable de la reducción total sufrimiento que deseamos. Además, el compromiso significa que los reductores de sufrimiento también pueden salir ganando en el escenario (con un ~35% de probabilidad, en mi opinión) en el que el desarrollo tecnológico de hecho prevenga de hecho más sufrimiento del que causa, incluso dejando de lado las consideraciones relativas a los compromisos estratégicos con otras personas.

Lo ideal sería que estos compromisos tomaran la forma de acuerdos negociados sólidos. Algunos ejemplos son posibles incluso a corto plazo. Por ejemplo, si los reductores de sufrimiento y los defensores de la colonización espacial se comprometen a cancelar la financiación de proyectos que entren en conflicto y, en vez de esto, desarrollan algún proyecto acordado en común.

La cuestión estratégica de dónde invertir los recursos para avanzar los valores que tenemos en un momento dado equivale a un dilema del prisionero con otros sistemas de valores. Y como a menudo tomamos decisiones acerca de dónde invertir, qué posturas adoptar y qué políticas impulsar, estos dilemas del prisionero son iterados. En [los torneos de Robert Axelrod](#) sobre el dilema del prisionero iterado, las estrategias más exitosas fueron siempre las consistentes en comportarse bien, es decir, en no ser el primero en abandonar al otro u otra. Por lo tanto, los reductores de sufrimiento no deben ser los primeros en traicionar a los colonizadores espaciales. Por supuesto, si parece que los colonizadores espaciales no están haciendo nada para reducir el sufrimiento, debemos también tener la disposición para que esto nos "provoque" a dejar temporalmente de cumplir nuestros acuerdos con ellos, hasta que el otro lado empiece a reconocer nuestras preocupaciones.

Quienes nos sentimos intranquilos e intranquilas ante la colonización del espacio, tenemos mucho que ganar si nos aliamos con sus partidarios –a la hora de pensar en los diversos escenarios que podrían darse, y de dar forma a un futuro en mejores direcciones–. También queremos mantener una relación amistosa con las personas a favor

de la colonización porque esto significa que estas tomarán nuestras ideas más en serio. Incluso si la colonización espacial ocurre, quedarán muchas sub-cuestiones sobre las que los reductores de sufrimiento quieren tener voz y voto: por ejemplo, su oposición a la difusión de la vida salvaje, a la creación de simulaciones/subrutinas que sufran, etc.

Queremos asegurarnos de que los reductores de sufrimiento no se conviertan en un grupo despreciado. Pensemos, por ejemplo, en cómo la eugenesia es un tema más tabú debido a las atrocidades nazis de lo que hubiera sido de otra manera. La gente anti-tecnología a veces es equiparada con el Unabomber. La reputación de los defensores y defensoras de los animales pueden ser empañadas por las tácticas violentas de unos pocos, o incluso por los comportamientos ridículos de PETA. Tenemos que tener cuidado de que no les ocurra algo similar a quienes quieren reducir el sufrimiento. A la mayoría de la gente ya le importa mucho la prevención del sufrimiento, y no queremos que la gente empiece a decir: "Ah, ¿te preocupas por la prevención del daño a criaturas indefensas? ¿Qué pasa, que eres uno de esos *reductores de sufrimiento*?", en un contexto en el que los llamados "reductores de sufrimiento" han cogido tan mala fama, que su nombre provoca automáticamente una reacción de hostilidad.

Así que no solo la cooperación con los partidarios de la colonización es la opción más prometedora, sino que es sin duda la única opción netamente positiva para nosotros. Adoptar una postura de mayor confrontación corre el riesgo de endurecer a la oposición y causar rechazo hacia nuestro mensaje. Recordemos que la prevención del sufrimiento futuro es algo que le importa a todo el mundo, y no debemos erosionar este hecho por ser excesivamente antagónicos.

9. Posibles aspectos positivos de un futuro inteligente

9.1. Cisnes negros que no sean armas de doble filo

Muchos escenarios especulativos que permitirían reducir enormemente el sufrimiento en el multiverso también permitirían aumentarlo enormemente: cuando se puede reducir el número de organismos que existen, también se puede aumentar dicho número, y los que

favorecen la creación de más felicidad/vida/complejidad/etc. tenderán a querer a apoyar el crecimiento.

Sin embargo, puede haber algunos cisnes negros que realmente no sean armas de doble filo, en el sentido de que más conocimiento es más probable que resulte en una disminución del sufrimiento. Por ejemplo, podríamos descubrir que algunas de nuestras rutinas físicas encajan en nuestras concepciones del sufrimiento. Y se podrían desarrollar formas de rediseñar esos procesos físicos para evitar el sufrimiento que contienen. Si esto se pudiese hacer sin un gran sacrificio de la felicidad o de otros valores, a la mayoría de la gente le parecería bien, suponiendo que los valores de hoy sigan teniendo algo de cuota de representación en decisiones futuras.

Esto puede ser algo bastante importante. Le asigno una probabilidad no trivial (¿tal vez ~10%?) a que, tras una reflexión pausada, yo adoptaría una visión altamente inclusiva de lo que cuenta como sufrimiento, de modo que sentiría que una parte significativa de todo el universo contiene procesos físicos de denso sufrimiento. Después de todo, la mecánica del sufrimiento puede ser vista como algo muy simple cuando se piensa sobre ella en cierta manera. Y lo que a mi parecer puede hacer que el sufrimiento animal resulte especial son solo ciertos atributos físicos peculiares (no necesariamente esenciales para la capacidad de sufrir) que poseen los animales (aunque podrían poseer otras). No está claro por qué estas características físicas particulares, y no otras, tendrían que ser más relevantes moralmente que "el número de patas, la velloidad de la piel, o la terminación del hueso sacro" ([citando a](#) Bentham). ¿Por qué no pueden otros objetos físicos en el universo constituir atributos físicos no esenciales para la posesión de sufrimiento que no obstante estarían presentes junto a procesos físicos que en cambio sí que darían lugar al sufrimiento? Todo esto es muy especulativo, pero en este momento sólo podemos comenzar a imaginar las formas de entender el universo a las que llegarán nuestros descendientes.

9.2. Valorando la reflexión

Si nos preocupamos en cierta medida por la reflexión moral sobre nuestros propios valores, no se sigue que la reducción del sufrimiento de un tipo particular sea sin duda el mejor camino. Por el contrario, ello es algo que nos da razones de más peso para apoyar un futuro de tecnología avanzada, por lo menos *si* es reflexivo.

En un escenario ideal, como por ejemplo el de la [volición coherente extrapolada](#) (CEV), si la reducción del sufrimiento fuese la visión moral más convincente los demás también se darían cuenta de ello.^[2] De hecho, en tal escenario se pondrían sobre la mesa para su consideración todos los argumentos que desarrollase cualquier filósofo moral (y muchos más que ningún filósofo hubiese desarrollado hasta entonces). Y la gente tendría la oportunidad de experimentar incluso un sufrimiento extremo, de manera controlada, a fin de evaluar lo malo que es comparado con otras cosas. Tal vez habría enfoques analíticos para predecir lo que la gente podría decir acerca de lo mala que es la tortura sin llegar a torturar a nadie para averiguarlo. Y, por supuesto, podríamos leer el registro histórico de la humanidad, y todos los textos sobre esto en internet para aprender más acerca de lo que la gente ha dicho sobre la tortura, a pesar de que necesitaríamos corregir los cálculos teniendo en cuenta el sesgo del "deseo de vivir", y los déficits de exactitud al recordar emociones en retrospectiva. Pero lo que relevante aquí es que en un escenario de CEV, todas estas consideraciones en el análisis de la cuestión pueden ser examinadas por gente mucho más inteligente que nosotros.

Por supuesto esta visión de color de rosa *no* es necesariamente un resultado futuro probable. Históricamente, quienes tienen el control de una cierta situación son quienes tienen más éxito a la hora de imponer su poder. Es bastante plausible que alguien se imponga en el futuro haciendo caso omiso de los deseos de todos los demás, y que lo que haga no sea buscar lo que para estos distintos deseos sea ideal e intentar conseguir una combinación de estos. O tal vez la preocupación por los débiles se quede en el camino, porque no sea muy adaptativo que los agentes poderosos se preocupen por los débiles. Esto sugiere que la mejora de las perspectivas de un futuro reflexivo y tolerante puede ser una tarea importante. En vez de centrarse en si el futuro ocurre, creo que es más valioso para quienes buscan reducir el sufrimiento el centrarse en conseguir que, *si* el

futuro sucede, este sea mejor –fomentando los acuerdos, la reflexión moral, la sabiduría filosófica, y el altruismo, los cuáles es esperable que mejoren la situación de todos–.

10. Agradecimientos

Este ensayo fue inspirado a raíz de una pregunta de Rob Wiblin. Carl Shulman me animó a discutir la cuestión de la cooperación con otros sistemas de valores. Al principio me resistí a aceptar su razonamiento, pero –como ha sucedido a menudo, al menos en lo que atañe a cuestiones de hecho y estratégicas– al final me di cuenta de que tenía razón, y acepté su punto de vista.

11. Notas a pie de página

[1] Debido a que la naturaleza contiene [cantidades de sufrimiento tan enormes](#), me desagradaría mucho tal proyecto. Incluyo esta cita por su efecto retórico y para dar una idea de cómo ven la situación otros.

[2] Por supuesto, lo que es atractivo para "mi yo idealizado" no es necesariamente atractivo para "tu yo idealizado". Pueden mantenerse divergencias de valores tras la extrapolación que se lleve a cabo. Dicho esto, no hay mejor alternativa que comprometernos con un enfoque como el de CEV, porque si trato de abandonar el acuerdo y favorecer mis valores particulares, entonces tú lo que harás será simplemente tratar de favorecer los tuyos, y es esperable que al final ambos saldremos perdiendo.

Inicio: [Ensayos sobre la reducción del sufrimiento](#)