

ph

LA MATRIX COMO METAFÍSICA*

DAVID CHALMERS

UNIVERSITY OF ARIZONA

RESUMEN ABSTRACT

El artículo intenta defender la tesis según la cual el escenario que se describe en *The Matrix* no corresponde a ninguna hipótesis escéptica. Antes bien, las experiencias que se viven en el interior de la matrix pueden ser tan reales como las que se dan por fuera de ella. Asimismo, y aunque el mundo de la matrix es completamente simulado, las experiencias que se viven en su interior no tienen por qué ser completamente falsas e incluso pueden ser verdaderas. Para desarrollar estas ideas, se exponen varias hipótesis que ayudan a clarificar el problema. De igual forma, se responden las objeciones que se han planteado contra estas tesis.

This paper tries to defend the thesis that the scenario described in *The Matrix* does not correspond to any skeptical hypothesis. Quite the contrary, the experiences that one has inside the matrix can be as real as those one has in the external world. Furthermore, though the matrix' world is totally simulated, the experiences one has inside it do not have to be entirely false, they may be even true. To argue for these ideas, several hypotheses are examined. Finally, the objections raised against these theses are answered.

PALABRAS CLAVE KEY WORDS

Matrix, escepticismo, simulación computacional, mundo, creencias.

Matrix, skepticism, computational simulation, world, beliefs.

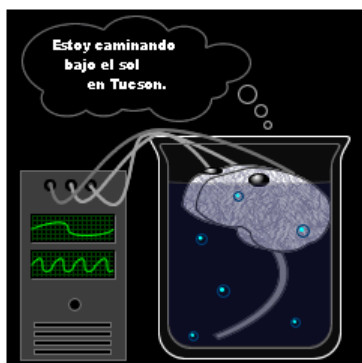
* Traducción del Inglés por Yobany Serna Castro

Recibido el 9 de Abril y aprobado el 30 de Julio de 2007.

I CEREBROS EN CUBETAS

The Matrix presenta una versión de una vieja fábula filosófica: el cerebro en una cubeta. Un cerebro separado de su cuerpo está flotando en una cubeta en el laboratorio de un científico. El científico ha dispuesto de antemano que el cerebro pueda ser estimulado con la misma clase de estímulos que recibe un cerebro normal en un cuerpo humano. Para hacer esto, el cerebro es conectado a una gigantesca simulación computarizada de un mundo. La simulación determina qué estímulos recibe el cerebro. Cuando el cerebro produce respuestas, éstas se retroalimentan dentro de la simulación. El estado interno del cerebro es exactamente igual como el de un cerebro normal, a pesar de que le falta un cuerpo. Desde el punto de vista del cerebro, las cosas son iguales a como nos aparecen a usted y a mí.

Al parecer, el cerebro está siendo sujeto a un engaño masivo. Tiene toda clase de creencias falsas sobre el mundo. Cree que tiene un cuerpo, pero no tiene ninguno. Cree que está caminando bajo la luz del sol, pero realmente está dentro de un laboratorio oscuro. Cree que está en un lugar, cuando de hecho puede estar en cualquier otro sitio completamente diferente. Quizás cree que está en Tucson, cuando realmente está en Australia, o incluso en el espacio exterior.



La situación de Neo al comienzo de *The Matrix* es parecida a esto. Él piensa que vive en una ciudad, piensa que tiene cabello, piensa que es 1999, y piensa que en la calle hay un sol radiante. En realidad, está flotando en el espacio, no tiene cabello, el año es alrededor del 2199, y el mundo ha sido oscurecido por la guerra. Hay algunas pequeñas diferencias del panorama de la cubeta mencionado anteriormente: el cerebro

de Neo está ubicado en un cuerpo, y la simulación computarizada está controlada por las máquinas en vez de un científico. Pero los detalles esenciales son en su mayoría los mismos. En efecto, Neo es un cerebro en una cubeta.

Vamos a decir que una *matrix* ("m" minúscula) es una simulación computarizada de un mundo diseñada artificialmente. Así, la Matrix

en la película es un caso particular de una matrix. Y vamos a decir que alguien está *encubetado*, o están *en una matrix*, si tiene un sistema cognitivo que recibe sus estímulos de, y envía sus respuestas a, la matrix. Así, el cerebro del ejemplo inicial está encubetado al igual que Neo.

Podemos imaginar que una matrix simula la física entera de un mundo, conservando hasta la última partícula a través del espacio-tiempo. (Más adelante miraremos el modo como podría variar esta estructura). Un ser encubetado podrá estar asociado con un cuerpo simulado particular. Se dispone una conexión de tal modo que cada vez que este cuerpo recibe estímulos sensoriales dentro de la simulación, el sistema cognitivo encubetado recibirá estímulos sensoriales del mismo tipo. Cuando el sistema cognitivo encubetado produce respuestas motoras, las respuestas correspondientes serán aplicadas a los órganos motores del cuerpo simulado.

Cuando se plantea la posibilidad de una matrix, inmediatamente surge una pregunta: ¿cómo sé que no estoy en una matrix? Después de todo, podría haber un cerebro en una cubeta estructurado exactamente como mi cerebro, conectado a una matrix, con experiencias indistinguibles de las que estoy teniendo ahora. Desde el interior, no hay ninguna manera de decir con seguridad que no estoy en la situación del cerebro en una cubeta. Parece, entonces, que no hay manera de saber con seguridad que no estoy en una matrix.

Denominemos a la hipótesis de que estoy en una matrix y que siempre he estado en una matrix la *Hipótesis de la Matrix*. De manera equivalente, la Hipótesis de la Matrix dice que estoy encubetado y he estado siempre encubetado. Esto no es completamente equivalente a la hipótesis de que estoy en la Matrix, en la medida en que la Matrix es sólo una versión específica de una matrix. Por ahora, ignoraré algunas de las complicaciones que son específicas de la Matrix de la película, tal como el hecho de que la gente algunas veces viaja de adentro hacia afuera entre la Matrix y el mundo externo. Dejando estos problemas a un lado, podemos interpretar informalmente la Hipótesis de la Matrix como si afirmara que estoy en la misma clase de situación que las personas que siempre han estado en la Matrix.

Deberíamos tomar en serio la Hipótesis de la Matrix. Como Nick Bostrom ha sugerido, no es inadmisible que en la historia del universo, la tecnología se desarrollará de tal modo que se puedan crear simulaciones

computarizadas de mundos enteros. Podría haber un vasto número de tales simulaciones, comparadas sólo con un único mundo real. Si es así, puede haber muchos más seres que están en una matrix que seres que no están. Dado todo esto, uno puede inferir incluso que es más probable que estemos en una matrix a que no estemos. Si esto es correcto o no, ciertamente parece que no podemos estar *seguros* de que no estamos en una matrix.

Parecen seguirse graves consecuencias. Mi contraparte (*counterpart*)¹ encubetada parece estar siendo masivamente engañada. Piensa que está en Tucson, que está sentado ante un escritorio escribiendo un artículo; piensa que tiene un cuerpo. Pero teniendo en cuenta esta situación, todas estas creencias son falsas. Asimismo, parece que si estoy encubetado, mis creencias correspondientes son falsas. Si estoy encubetado, no estoy realmente en Tucson, realmente no estoy sentado ante un escritorio, y podría no tener un cuerpo siquiera. Por lo tanto, si no sé que no estoy encubetado, entonces no sé que estoy en Tucson, no sé que estoy sentado ante un escritorio, y no sé que tengo un cuerpo.

La Hipótesis de la Matrix amenaza con socavar casi todo lo que sé. Parece ser una *hipótesis escéptica*: una hipótesis que no puedo eliminar, y que si fuera verdadera, falsaría la mayoría de mis creencias. Allí donde hay una hipótesis escéptica, parece como si ninguna de las creencias respectivas constituyera conocimiento genuino. Por supuesto, las creencias *podrían* ser verdaderas —puedo ser afortunado y no estar encubetado— pero no puedo eliminar la posibilidad de que sean falsas. De esta manera, una hipótesis escéptica conduce al *escepticismo* sobre estas creencias: creo estas cosas, pero no las conozco.

Para resumir el razonamiento: no puedo saber que no estoy en una matrix. Si estoy en una matrix, probablemente no estoy en Tucson. Si no sé que no estoy en una matrix, entonces no puedo saber que estoy en Tucson. Lo mismo se aplica para casi todo lo que creo conocer sobre el mundo externo.

¹ Esta expresión fue introducida por David Lewis, quien defiende la tesis de que el nuestro es sólo uno de entre muchos posibles mundos, todos igualmente reales. Y que en otros de esos mundos existen individuos muy similares a cada uno de nosotros, los cuales son nuestras contrapartes (*counterparts*). (*N del T*).

II UNA RECONSIDERACIÓN DEL CEREBRO ENCUBETADO

Esta es una forma estándar de pensar acerca de la situación de la cubeta. Parece que este punto de vista también es aceptado por quienes crearon *The Matrix*. En el caso de la película para DVD, uno ve lo siguiente:

Percepción: nuestro mundo interno-externo es real.

Realidad: tal mundo es una broma, un engaño elaborado e hilado por las máquinas todopoderosas que nos controlan.

Pienso que este punto de vista no es completamente correcto. Pienso que, incluso si estoy en una matrix, mi mundo es perfectamente real. Un cerebro en una cubeta no tiene por qué estar sujeto a un engaño masivo (por lo menos si siempre ha estado en la cubeta). Neo no tiene creencias masivamente falsas sobre el mundo externo. Más bien, los seres encubetados tienen creencias ampliamente *correctas* sobre su mundo. Si esto es así, la Hipótesis de la Matrix no es una hipótesis escéptica, y su posibilidad no socava todo lo que creo conocer.

Los filósofos han sostenido este punto de vista antes. En el siglo XVIII el filósofo George Berkeley sostuvo, en efecto, que la apariencia es la realidad. (Recordemos a Morpheus: “¿Qué es real? ¿Cómo defines lo real? Si estás hablando sobre lo que puedes sentir, oler, gustar y ver, entonces lo real es simplemente una serie de señales eléctricas interpretadas por tu cerebro.”) Si esto es correcto, entonces el mundo percibido por seres encubetados es perfectamente real: ellos tienen todas las apariencias correctas, y la apariencia es la realidad. Desde este punto de vista, incluso los seres encubetados tienen creencias verdaderas sobre el mundo.

Me he encontrado asumiendo recientemente una conclusión similar, aunque por razones completamente diferentes. No encuentro plausible que la apariencia sea la realidad, por eso no apoyo el razonamiento de Berkeley. Y hasta hace poco me parecía bastante obvio que un cerebro encubetado tendría creencias masivamente falsas. Pero ahora pienso que hay una línea de razonamiento que muestra que esto es un error.

Aún pienso que no puedo descartar la hipótesis de que estoy en una matrix. Pero pienso que incluso si estoy en una matrix, estoy en Tucson, estoy sentado en mi escritorio, y así sucesivamente. Así que la hipótesis

de que estoy en una matrix no es una hipótesis escéptica. Lo mismo se aplica para Neo. Al comienzo de la película, si él piensa “tengo cabello”, está en lo correcto. Si piensa “el día es soleado”, está en lo correcto. Y lo mismo, por supuesto, para el cerebro original en una cubeta. Cuando éste piensa “tengo un cuerpo”, está en lo cierto. Cuando piensa “estoy caminando”, está en lo cierto.

Este punto de vista parece ser, en principio, bastante contraintuitivo. Inicialmente, me parecía completamente contraintuitivo. Pero presentaré ahora la línea de razonamiento que me ha convencido de que es correcto.

III LA HIPÓTESIS METAFÍSICA

Argumentaré que la hipótesis de que estoy encubetado no es una hipótesis escéptica, sino una *hipótesis metafísica*. Es decir, es una hipótesis sobre la naturaleza subyacente de la realidad.

Si la física se preocupa por los procesos microscópicos que subyacen a la realidad macroscópica, la metafísica se preocupa por la naturaleza fundamental de la realidad. Una hipótesis metafísica puede hacer una afirmación sobre la realidad que subyace a la física misma. Alternativamente, podría decir algo sobre la naturaleza de nuestras mentes, o la creación de nuestro mundo.

Opino que la Hipótesis de la Matrix debe ser considerada como una hipótesis metafísica con estos tres elementos: hace una afirmación sobre la realidad subyacente a la física, otra sobre la naturaleza de nuestras mentes, y una más sobre la creación del mundo.

En particular, opino que la Hipótesis de la Matrix es equivalente a una versión de la siguiente Hipótesis Metafísica tripartita. En primer lugar, los procesos físicos son fundamentalmente computacionales. En segundo lugar, nuestros sistemas cognitivos están separados de los procesos físicos, pero interactúan con éstos. En tercer lugar, la realidad física fue creada por seres fuera del espacio-tiempo físico.

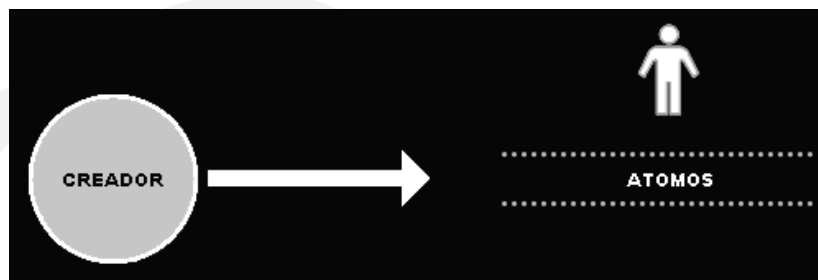
Es importante observar que nada en esta Hipótesis Metafísica es escéptico. La Hipótesis Metafísica nos habla aquí sobre los procesos subyacentes a nuestra realidad ordinaria, pero no implica que esta

realidad no existe. Todavía tenemos cuerpos, y hay aún sillas y mesas: es sólo que su naturaleza fundamental es un poco diferente de lo que podríamos haber pensado. De esta manera, la Hipótesis Metafísica es análoga a una hipótesis física, en la medida en que ésta tiene que ver con mecánica cuántica. Tanto la hipótesis física como la Hipótesis Metafísica nos hablan sobre los procesos subyacentes a las sillas. Ninguna implica que no existan sillas. Más bien, nos dicen cómo son realmente las sillas.

Construiré el argumento introduciendo por separado cada una de las partes de la Hipótesis Metafísica. Sugeriré que cada una de ellas es coherente, y no pueden ser conclusivamente descartadas. Y sugeriré que ninguna de ellas es una hipótesis escéptica: incluso si fueran verdaderas, la mayoría de nuestras creencias ordinarias serían correctas. Lo mismo se aplica para una combinación de las tres hipótesis. Argumentaré, entonces, que la hipótesis que plantea la Hipótesis de la Matrix es equivalente a esta combinación.

(1) La Hipótesis de la Creación

La Hipótesis de la Creación dice: el espacio-tiempo físico y sus contenidos fueron creados por seres que están por fuera del espacio-tiempo físico.

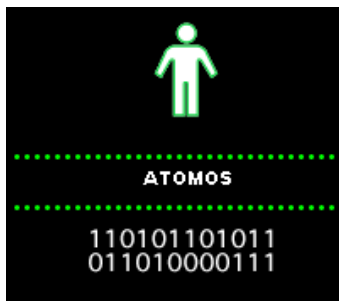


Esta es una hipótesis familiar. Una de sus versiones es aceptada por mucha gente en nuestra sociedad, y quizás por la mayoría de las personas en el mundo. Si uno cree que Dios creó el mundo, y si uno cree que Dios está fuera del espacio-tiempo físico, entonces cree en la Hipótesis de la Creación. Sin embargo, uno no necesita creer en Dios para creer en la Hipótesis de la Creación. Quizá nuestro mundo fue creado por un ser relativamente ordinario en el “universo contiguo superior”, usando la tecnología más avanzada de ese universo para crear mundos. En este caso, la Hipótesis de la Creación es verdadera.

No puedo saber si la Hipótesis de la Creación es verdadera, pero no puedo saber con toda seguridad que es falsa. La hipótesis es claramente coherente, y no puedo hacerla a un lado de manera concluyente.

La Hipótesis de la Creación no es una hipótesis escéptica. Incluso si es verdadera, la mayoría de mis creencias ordinarias todavía son verdaderas. Todavía tengo manos, estoy en Tucson, y así sucesivamente. Quizás, algunas de mis creencias serán falsas: si soy un ateo, por ejemplo, o si creo que toda la realidad comenzó con el Big Bang. Pero la mayoría de mis creencias cotidianas sobre el mundo externo permanecerán intactas.

(2) La Hipótesis Computacional



La Hipótesis Computacional dice: los procesos microfísicos a lo largo del espacio-tiempo están constituidos por procesos computacionales subyacentes.

La Hipótesis Computacional afirma que la física, como nosotros la conocemos, no es el nivel fundamental de la realidad. Así como los procesos químicos sustentan los procesos biológicos, y los procesos microfísicos subyacen a los procesos químicos, algo subyace a los procesos microfísicos. Bajo el nivel de los quarks, los electrones y los fotones hay un nivel adicional: el nivel de los bits. Estos bits están gobernados por un algoritmo computacional, que a un nivel más alto produce los procesos que nosotros consideramos como partículas fundamentales, fuerzas, y así sucesivamente.

La Hipótesis Computacional no goza de una aceptación tan amplia como la Hipótesis de la Creación, pero algunas personas la toman en serio. Célebremente, Ed Fredkin postuló que el universo es en el fondo alguna clase de computadora. Más recientemente, Stephen Wolfram adoptó la idea en su libro *A New Kind of Science*, sugiriendo que en el nivel fundamental, la realidad física puede ser una clase de autómatas celulares, con interacción de bits gobernados por reglas simples. Y algunos físicos han considerado la posibilidad de que las leyes de la física puedan formularse computacionalmente, o puedan verse como la consecuencia de ciertos principios computacionales.

Uno podría sospechar de que meros bits sean el nivel fundamental de la realidad: un bit es sólo un 0 o un 1, y la realidad no puede consistir

en ceros y unos. O tal vez un bit es sólo una “pura diferencia” entre dos estados básicos, y no puede haber una realidad hecha de puras diferencias. Más bien, los bits siempre tienen que estar implantados en estados más básicos, como los voltajes en una computadora normal.

No sé si esta objeción es correcta. No creo que esté completamente fuera de lugar el que pueda haber un universo de “puros bits”. Pero esto no importa aquí. Podemos suponer que el nivel computacional mismo está constituido por un nivel más fundamental, en el cual los procesos computacionales son implementados. No importa para los propósitos del presente tema cuál es el nivel más fundamental. Todo lo que importa es que los procesos microfísicos están constituidos por procesos computacionales, los cuales, a su vez, están constituidos por procesos más básicos. Entenderé, de ahora en adelante, la Hipótesis Computacional de esta forma.

No sé si la Hipótesis Computacional es correcta. Pero tampoco sé si es falsa. Aunque especulativa, la hipótesis es coherente, y no puedo eliminarla definitivamente.

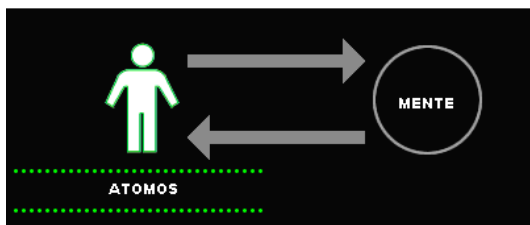
La Hipótesis Computacional no es una hipótesis escéptica. Si es verdadera, hay todavía electrones y protones. En esta imagen, los electrones y los protones pueden ser análogos a las moléculas: están hechos de algo más básico, pero todavía existen. De manera similar, si la Hipótesis Computacional es verdadera, todavía hay mesas y sillas, y la realidad macroscópica aún existe. Es solamente que su realidad fundamental es un poco diferente de lo que pensábamos.

La situación aquí es análoga a la que tiene que ver con la mecánica cuántica o la relatividad. Estas podrían conducirnos a revisar algunas creencias “metafísicas” acerca del mundo externo: como que el mundo está hecho de partículas clásicas, o que el tiempo es absoluto. Pero la mayoría de nuestras creencias ordinarias permanecen intactas. Igualmente, aceptar la Hipótesis Computacional puede llevarnos a revisar unas pocas creencias metafísicas: por ejemplo, que los electrones y los protones son fundamentales. Pero la mayoría de nuestras creencias ordinarias no son afectadas.

(3) La Hipótesis Mente-Cuerpo

La Hipótesis Mente-Cuerpo dice: mi mente está (y siempre ha estado) constituida por procesos fuera del espacio-tiempo físico, y recibe sus

estímulos perceptuales de, y envía sus respuestas a, los procesos del espacio-tiempo físico.



La Hipótesis Mente-Cuerpo es tan bien muy familiar y aceptada. Descartes sostuvo algo similar: en su opinión, tenemos mentes no físicas que interactúan con nuestros cuerpos físicos. La hipótesis

es hoy menos creída que en el tiempo de Descartes, pero hay todavía muchas personas que aceptan la Hipótesis Mente-Cuerpo.

Independientemente de si la Hipótesis Mente-Cuerpo es o no verdadera, ciertamente es coherente. Aún si la ciencia contemporánea tiende a sugerir que la hipótesis es falsa, no podemos descartarla concluyentemente.

La Hipótesis Mente-Cuerpo no es una hipótesis escéptica. Incluso si mi mente está fuera del espacio-tiempo físico, todavía tengo un cuerpo, todavía estoy en Tucson, y así sucesivamente. A lo sumo, aceptando esta hipótesis podemos revisar unas pocas creencias metafísicas acerca de nuestras mentes. Nuestras creencias ordinarias sobre la realidad externa permanecerán, en gran parte, intactas.

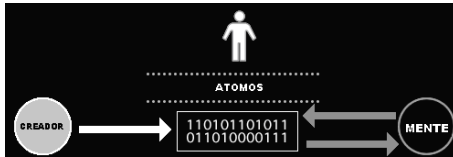
(4) La Hipótesis Metafísica

Podemos reunir ahora estas hipótesis. Podemos considerar la Hipótesis de la Combinación, la cual las une a todas. Esta hipótesis dice que el espacio-tiempo físico y sus contenidos fueron creados por seres fuera del espacio-tiempo físico. Que los procesos microfísicos están constituidos por procesos computacionales, y que nuestras mentes están fuera del espacio-tiempo físico pero interactúan con éste.

Como sucede con las hipótesis tomadas individualmente, la Hipótesis de la Combinación es coherente, y nosotros no podemos desecharla concluyentemente. Al igual que cada hipótesis individual, ésta no es una hipótesis escéptica. Aceptarla podría conducirnos a revisar algunas de nuestras creencias, pero dejaría intacta a la mayoría.

Finalmente, podemos considerar la Hipótesis Metafísica (con M mayúscula). Como la Hipótesis de la Combinación, ésta combina la Hipótesis de la Creación, la Hipótesis Computacional y la Hipótesis

Mente-Cuerpo. También agrega la siguiente afirmación más específica: los procesos computacionales que subyacen al espacio-tiempo físico, fueron diseñados por los creadores como una simulación computarizada de un mundo.



(También puede ser útil pensar en la Hipótesis Metafísica como si afirmara que los procesos computacionales que constituyen el espacio-tiempo físico son parte

de un dominio más amplio, y que los creadores y mi sistema cognitivo están también localizados dentro de este dominio. Esta adición no es estrictamente necesaria para lo que sigue, pero se relaciona con la manera más común de pensar acerca de la Hipótesis de la Matrix.)

La Hipótesis Metafísica es una versión un poco más específica de la Hipótesis de la Combinación, en cuanto que especifica algunas de las relaciones entre varias partes de las hipótesis. De nuevo, la Hipótesis Metafísica es una hipótesis coherente, y no podemos descartarla concluyentemente. Y tampoco es una hipótesis escéptica. Incluso si la aceptamos, la mayoría de nuestras creencias ordinarias sobre el mundo externo permanecerá intacta.

IV LA HIPÓTESIS DE LA MATRIX COMO UNA HIPÓTESIS METAFÍSICA

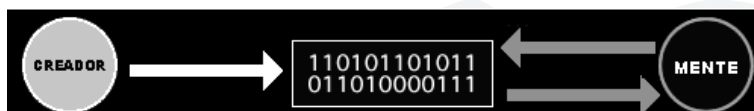
Recordemos que la Hipótesis de la Matrix dice: tengo (y siempre he tenido) un sistema cognitivo que recibe sus estímulos de, y envía sus respuestas a, una simulación computarizada, diseñada artificialmente, de un mundo.

Argumentaré que la Hipótesis de la Matrix es equivalente a la Hipótesis Metafísica en el siguiente sentido: si acepto la Hipótesis Metafísica, debería aceptar la Hipótesis de la Matrix, y si acepto la Hipótesis de la Matrix, debería aceptar la Hipótesis Metafísica. Esto es, las dos hipótesis se *implican* mutuamente, de donde se sigue que si uno acepta una, debería aceptar la otra.

Sigamos la primera vía desde la Hipótesis Metafísica a la Hipótesis de la Matrix. La Hipótesis Mente-Cuerpo implica que tengo (y siempre

he tenido) un sistema cognitivo aislado que recibe sus estímulos de, y envía sus respuestas a, procesos en el espacio-tiempo físico. En conjunción con la Hipótesis Computacional, esto implica que mi sistema cognitivo aislado recibe estímulos de, y envía respuestas a, los procesos computacionales que constituyen el espacio-tiempo físico. La Hipótesis de la Creación (junto con el resto de la Hipótesis Metafísica) implica que estos procesos fueron diseñados artificialmente para simular un mundo. Se sigue que tengo (y siempre he tenido) un sistema cognitivo aislado que recibe sus estímulos de, y envía sus respuestas a, una simulación artificialmente diseñada de un mundo. Esta es precisamente la Hipótesis de la Matrix. Así que la Hipótesis Metafísica implica la Hipótesis de la Matrix.

La otra vía es muy similar. Para decirlo de manera informal: si acepto la Hipótesis de la Matrix, acepto que lo que subyace a la realidad aparente es algo como lo que especifica la Hipótesis Metafísica. Existe un dominio que contiene mi sistema cognitivo, que interactúa causalmente con una simulación computarizada del espacio-tiempo físico, que fue creado por otros seres en ese dominio. Esto es lo que hay que tener para obtener la Hipótesis Metafísica. Si uno acepta esto, podría aceptar la Hipótesis de la Creación, la Hipótesis Computacional, la Hipótesis Mente-Cuerpo, y las relaciones relevantes entre éstas.



Esto puede ser un poco más claro a través de la imagen. Aquí está la forma del mundo según la Hipótesis de la Matrix.

En el nivel fundamental, esta imagen de la forma del mundo es exactamente igual a la imagen de la Hipótesis Metafísica que mostré atrás. Así que si uno acepta que el mundo es como en la Hipótesis de la Matrix, podría aceptar que es como en la Hipótesis Metafísica.

Podrían plantearse varias objeciones. Por ejemplo, podría objetarse que la Hipótesis de la Matrix implica que existe una simulación computarizada de los procesos físicos, pero (al contrario de la Hipótesis Metafísica) esto no implica que los procesos físicos en sí mismos existan. Discutiré esta objeción en la sección 6, y otras objeciones en la sección 7. Por ahora, sin embargo, asumiré que hay una buena razón para pensar que la Hipótesis de la Matrix implica la Hipótesis Metafísica, y viceversa.

V LA VIDA EN LA MATRIX

Si esto es correcto, se sigue que la Hipótesis de la Matrix no es una hipótesis escéptica, porque si la acepto no podría inferir que el mundo externo no existe, o que no tengo un cuerpo, o que no hay mesas y sillas, o que no estoy en Tucson. Más bien, debo inferir que el mundo físico está constituido por cálculos que subyacen al nivel microfísico. Todavía existen mesas, sillas y cuerpos: éstos se componen fundamentalmente de bits, y de cualquier cosa de la que estén hechos los bits. Este mundo fue creado por otros seres, pero aún así es perfectamente real. Mi mente está separada de los procesos físicos e interactúa con éstos. Mi mente podría no haber sido creada por estos seres, y podría no estar compuesta por estos bits, pero todavía interactúa con éstos.

El resultado es una imagen compleja de la naturaleza fundamental de la realidad. La imagen, quizás, es extraña y sorprendente, pero es una imagen de un mundo externo de carne y hueso. Si estamos en una matrix, esta simplemente es la forma en que es el mundo.

Podemos pensar en la Hipótesis de la Matrix como un mito de la creación para la era de la información. Si esto es correcto, entonces el mundo físico fue creado, pero no necesariamente por dioses. Por debajo del mundo físico hay un cálculo gigante, y los creadores crearon este mundo mediante la implementación de este cálculo. Y nuestras mentes existen fuera de esta estructura física, con una naturaleza independiente que interactúa con esta estructura.

Muchos de los mismos problemas que se plantean respecto de los mitos creacionistas estándar surgen aquí. ¿Cuándo fue creado el mundo? Estrictamente hablando, no se creó dentro de *nuestro* tiempo. ¿Cuándo empezó la historia? Los creadores podrían haber empezado la simulación en el 4004 antes de Cristo (o en 1999) con el registro fósil intacto, pero habría sido mucho más fácil para ellos haber iniciado la simulación con el Big Bang y permitir que las cosas siguieran su curso normal desde ahí.

(En la película *Matrix*, desde luego, los creadores son máquinas. Esto ofrece un interesante giro sobre las interpretaciones teológicas comunes de la película. A menudo se sostiene que Neo es la figura de Cristo en la película, que Morpheus corresponde a Juan el Bautista, Cypher a Judas Iscariote, y así sucesivamente. Pero en la interpretación que he

hecho, los dioses de la Matrix son las máquinas. Entonces, ¿quién es la figura de Cristo? ¡El agente Smith, por supuesto! Después de todo, él es la descendencia de los dioses, enviado a salvar el mundo de la Matrix de aquellos que desean destruirlo. Y en la segunda película, incluso él resucita.)

Muchos de los mismos problemas que surgen con la Hipótesis Mente-Cuerpo también surgen aquí. ¿Desde cuándo nuestras mentes no-físicas empezaron a existir? Simplemente depende de cuándo un nuevo sistema cognitivo encubetado está atado a la simulación (¿Quizás en el momento de la concepción dentro de la matrix, o quizás en el momento del nacimiento?) ¿Existe vida después de la muerte? Depende simplemente de lo que le pase a los sistemas encubetados una vez sus cuerpos simulados mueran. ¿Cómo interactúan la mente y el cuerpo? Por conexiones causales que están fuera del espacio-tiempo físico.

Incluso si no estamos en una matrix, podemos aplicar una versión de este razonamiento a otros seres que están en una matrix. Si descubren su situación, y aceptan que están en una matrix, no deben rechazar sus creencias ordinarias sobre el mundo externo. A lo sumo, deben revisar sus creencias sobre la naturaleza subyacente de su mundo: deberían aceptar que los objetos externos están hechos de bits, y así sucesivamente. Estos seres no son masivamente engañados: la mayoría de sus creencias ordinarias sobre su mundo son correctas.

Aquí deben introducirse algunas aclaraciones. Uno podría sospechar de las creencias sobre las otras mentes. Creo que mis amigos son conscientes. Si estoy en una matrix, ¿esto es correcto? En la Matrix descrita en la película esta creencia es mayoritariamente correcta. Esta es una matrix multi-cubeta: para cada uno de mis amigos percibidos, hay un ser encubetado que está en la realidad externa, quien es presumiblemente consciente como yo. La excepción podría ser un ser como el agente Smith, quien no está encubetado, sino que es completamente computacional. El que estos seres sean conscientes, depende de si el cómputo es suficiente para tener conciencia. Permaneceré neutral sobre este problema. Podríamos evitar este problema introduciendo en la Hipótesis de la Matrix el requisito de que todos los seres que percibimos estén encubetados. Pero aún si nos basamos en este requerimiento, no estamos mucho peor que en el mundo real donde existe un problema legítimo acerca de si otros seres son conscientes, independientemente de si estamos en una matrix o no.

Uno también podría dudar de las creencias sobre el pasado distante y sobre el futuro lejano. Éstas permanecerán intactas con tal de que la simulación computarizada cubra todo el espacio-tiempo, desde el Big Bang hasta el final del universo. Esto hace parte de la Hipótesis Metafísica, y podemos estipular, además, que hace parte de la Hipótesis de la Matrix también, exigiendo que la simulación computarizada sea una simulación de un mundo entero. Puede haber otras simulaciones que comiencen en el pasado reciente (quizás la Matrix en la película es como esto) y puede haber otras que sólo existen por un periodo muy corto. En estos casos, los seres encubetados tendrán creencias falsas sobre el pasado y/o el futuro en sus mundos. Pero mientras la simulación cubra toda la vida de estos seres, es plausible que tendrán la mayoría de las veces creencias correctas sobre el estado actual de su entorno.

Puede haber algunos temas en los cuales los seres en una matrix son engañados. Puede ser que los creadores de la matrix controlan, e interfieren con, mucho de lo que sucede en el mundo simulado. (La Matrix en la película puede ser como esto, aunque la extensión del control de los creadores no es lo suficientemente clara). En este caso, estos seres pueden tener mucho menos control sobre lo que sucede de lo que ellos mismos piensan. Pero sucede lo mismo si un dios interfiere en un mundo no-matrix. Y la Hipótesis de la Matrix no implica que los creadores interfieran con el mundo, aunque se deja abierta esta posibilidad. En el peor de los casos, la Hipótesis de la Matrix no es más escéptica en este respecto que la Hipótesis de la Creación en un mundo no-matrix.

Los habitantes de una matrix también pueden ser engañados en que la realidad es mucho más grande de lo que ellos piensan. Pueden pensar que su universo físico es todo lo que hay, cuando de hecho hay mucho más en el mundo, incluyendo seres y objetos que posiblemente nunca puedan ver. Pero de nuevo, esta clase de preocupación puede surgir igualmente en un mundo no-matrix. Por ejemplo, los cosmólogos consideran seriamente la hipótesis de que nuestro universo puede provenir de un agujero negro en un "universo contiguo superior", y que en realidad puede haber todo un árbol de universos. En este caso, el mundo es mucho más grande de lo que pensamos, y puede haber seres y objetos que posiblemente no podremos ver. Pero de cualquier modo, el mundo que vemos es perfectamente real.

De manera importante, ninguna de estas fuentes de escepticismo - sobre otras mentes, el pasado y el futuro, sobre nuestro control de otros

mundos, y sobre la extensión del mundo- plantea dudas sobre nuestra creencia en la realidad del mundo que percibimos. Ninguna de ellas nos lleva a dudar de la existencia de objetos externos tales como mesas y sillas, en la forma en que se supone que lo hace la hipótesis de la cubeta. Y ninguna de estas preocupaciones se ajusta especialmente a la situación de la matrix. Uno puede plantear dudas sobre si otras mentes existen, si el pasado y el futuro existen, y si tenemos el control suficiente sobre nuestros mundos, independientemente de si estamos en una matrix o no. Si esto es correcto, entonces la Hipótesis de la Matrix no plantea los problemas escépticos distintivos que a menudo se asume que plantea.

Sugerí antes que no está fuera de toda duda el que estemos realmente en una matrix. Podría pensarse que esto es una conclusión preocupante. Pero si tengo razón, esto no es tan preocupante como se podría haber pensado. Incluso si estamos en una matrix, nuestro mundo no es menos real de lo que pensamos que era. Apenas tiene una naturaleza fundamental sorprendente.

VI OBJECIÓN: LA SIMULACIÓN NO ES LA REALIDAD (Esta sección un poco técnica puede pasarse por alto sin mucha pérdida.)

Una línea común de objeción es que una simulación no es lo mismo que la realidad. La Hipótesis de la Matrix sólo implica que existe una simulación de procesos físicos. Por contraste, la Hipótesis Metafísica implica que realmente existen los procesos físicos (ellos son explícitamente mencionados en la Hipótesis Computacional y en otro lugar). En este caso, entonces la Hipótesis de la Matrix no puede implicar la Hipótesis Metafísica. Desde este punto de vista, si estoy en una matrix, entonces los procesos físicos no existen realmente.

Respuesta: mi argumento no requiere la suposición general de que la simulación es lo mismo que la realidad. El argumento opera de una manera bastante diferente. Pero la objeción nos ayuda a construir el argumento informal de que la Hipótesis de la Matrix implica la Hipótesis Metafísica.

Como la Hipótesis Computacional es coherente, es claramente *posible* que un nivel computacional subyazca a los procesos físicos reales, y es posible que, a su vez, los cómputos sean implementados por procesos

adicionales. De esta manera, existe *alguna* clase de sistema computacional que puede fundamentar la realidad. Pero en este punto el oponente replicará que no todos los sistemas computacionales son creados de igual forma. Decir que algunos sistemas computacionales fundamentarán procesos físicos reales en este rol, no es lo mismo que decir que todos lo hacen. Posiblemente, algunos de éstos son meramente simulaciones. En este caso, entonces la Hipótesis de la Matrix no implica la realidad.

Para refutar esta objeción, podemos apelar a dos principios. Primer principio: cualquier cómputo abstracto que puede ser usado para simular el espacio-tiempo físico es tal que *podría* convertirse en una base de los procesos físicos reales. Segundo principio: dado un cómputo abstracto que *puede* sustentar los procesos físicos, la forma precisa como sea implementado es irrelevante para la cuestión de si subyace a los procesos físicos. En particular, el hecho de que la implementación fuera diseñada como una simulación, es irrelevante. La conclusión, entonces, se sigue directamente.

Acerca del primer principio: pensemos en los cálculos abstractos en términos puramente formales, abstrayéndolos de sus modos de implementación. Para que un cálculo abstracto sea una simulación de la realidad física, debe tener elementos computacionales que correspondan a cada partícula en la realidad (igualmente para los campos, ondas, o cualquier otra cosa que sea fundamental), evolucionando dinámicamente de manera que corresponda a la evolución de las partículas. Pero entonces, se garantiza que el cálculo pueda tener una estructura causal lo suficientemente rica para que *pueda*, en principio, sustentar la física de nuestro mundo. Cualquier cálculo servirá en la medida en que tenga los suficientes detalles para corresponder a los finos detalles de los procesos físicos.

Acerca del segundo principio: dado un cálculo abstracto que pueda sustentar la realidad física, no importa cómo el cálculo sea implementado. Podemos imaginarnos el descubrimiento de algunos niveles computacionales que sustentan el nivel de los átomos y los electrones. Una vez hayamos descubierto esto, es posible que este nivel computacional esté implementado por procesos más básicos. Hay muchas hipótesis sobre cómo los procesos subyacentes podrían ser, pero ninguna de éstas es especialmente privilegiada, y ninguna puede hacernos rechazar la hipótesis de que el nivel computacional constituye los procesos físicos. Esto es, la Hipótesis Computacional es *independiente*

de su implementación: con tal de que tengamos la clase correcta de cómputo abstracto, la manera de implementarlo no importa.

En particular, es irrelevante si estos procesos de implementación fueron creados artificialmente o no, y es irrelevante si fueron pensados como una simulación. Lo que importa es la naturaleza intrínseca de los procesos, no su origen. Y lo que importa sobre esta naturaleza intrínseca es simplemente que ellos están constituidos de manera tal que implementan la clase correcta de cómputo. Si esto es así, el hecho de que la implementación sea originada como una simulación es irrelevante para la cuestión de si puede constituir la realidad física.

Hay una restricción adicional sobre los procesos implementadores: ellos deben estar, de esta manera, conectados a nuestras experiencias de la forma correcta. Esto es, cuando tenemos una experiencia de un objeto, los procesos subyacentes a la simulación de ese objeto deben estar causalmente conectados de la forma correcta con nuestras experiencias. Si este no es el caso, entonces no habrá ninguna razón para pensar que estos procesos computacionales subyacen a los procesos físicos que percibimos. Si hay una simulación computarizada aislada, a la cual nadie está conectado de esta manera, podemos decir que es simplemente una simulación. Pero en la Hipótesis de la Matrix, hay una conexión apropiada con nuestras experiencias preceptuales, por lo menos en la interpretación más natural de esa hipótesis. Así, la Hipótesis de la Matrix no tiene ningún problema aquí.

En general, entonces, hemos visto que un proceso computacional *puede* subyacer a la realidad física, que un cómputo abstracto que sea una simulación de la realidad física puede jugar este papel, y que una implementación de este cómputo puede constituir la realidad física, con tal de que esté conectado a nuestras experiencias de la manera relevante. La Hipótesis de la Matrix garantiza que tenemos un cómputo abstracto de la clase correcta, y que éste está conectado a nuestras experiencias del modo relevante. Entonces, la Hipótesis de la Matrix implica que la Hipótesis Computacional es correcta, y que la simulación computarizada constituye procesos físicos genuinos.

VII OTRAS OBJECIONES

Cuando miramos un cerebro en una cubeta desde el exterior, es difícil evitar el sentimiento de que está siendo engañado. Este mismo sentimiento se manifiesta en varias objeciones relacionadas. Éstas no son objeciones directas al argumento anterior, pero son objeciones a su conclusión.



Objeción 1: un cerebro en una cubeta puede pensar que está en el exterior caminando bajo el sol, cuando de hecho está solo en una habitación oscura. ¡Ciertamente es engañado!

Respuesta: el *cerebro* está solo en una habitación oscura. Pero esto no implica que la *persona* esté sola en una habitación oscura. Por analogía, digamos que Descartes está en lo cierto en cuanto piensa que tenemos

mentes no corporales fuera del espacio-tiempo, hechas de ectoplasma. Cuando pienso "estoy afuera bajo el sol", un ángel podría mirar mi mente ectoplásmica y notar que de hecho no estoy expuesto en absoluto a ningún sol. ¿De esto se sigue que mi pensamiento es incorrecto? Probablemente no: puedo estar afuera bajo el sol, incluso si mi mente ectoplásmica no lo está. El ángel puede estar equivocado al inferir que tengo una creencia incorrecta. Igualmente, no podemos inferir que el ser encubetado tiene una creencia incorrecta. Por lo menos, no está más engañado que una mente cartesiana.

La moraleja es que los ambientes inmediatos de nuestras mentes podrían ser irrelevantes con respecto a la verdad de la mayoría de nuestras creencias. Lo que importa son los procesos a los que nuestras mentes están conectadas, por estímulos perceptuales y respuestas motoras. Una vez reconozcamos esto, la objeción desaparecerá.

Objeción 2: un ser encubetado puede creer que está en Tucson, cuando de hecho está en New York, y nunca ha estado en alguna parte cerca de Tucson. Ciertamente esta creencia es engañosa.

Respuesta: el concepto de "Tucson" del ser encubetado no se refiere a lo que llamamos Tucson. Más bien, se refiere a algo completamente distinto: llamémoslo Tucson*, o "Tucson virtual". Podríamos pensar en

esto como una “ubicación virtual” (diré más sobre esto en un momento). Cuando el ser se dice a sí mismo “estoy en Tucson”, realmente está pensando que está en Tucson*, y podría estar de hecho en Tucson'. Porque Tucson no es Tucson*, el hecho de que el ser nunca haya estado en Tucson es irrelevante con respecto a si su creencia es verdadera.

Una analogía simple: miro a mi colega Terry, y pienso “ése es Terry”. En otra parte del mundo, un duplicado de mí mira un duplicado de Terry. Piensa “ése es Terry”, pero no está mirando al Terry real. ¿Es su creencia falsa? Parece que no: el concepto de “Terry” de mi duplicado no se refiere a “Terry”, sino a su Terry* duplicado. Mi duplicado realmente está mirando a Terry*, por ello su creencia es verdadera. La misma clase de cosa está pasando en el caso anterior.

Objeción 3: antes de dejar la Matrix, Neo cree que tiene cabello. Pero en realidad no tiene cabello (el cuerpo en la cubeta es calvo). Ciertamente su creencia es engañosa.

Respuesta: este caso es como el último. El concepto de Neo de “cabello” no se refiere al cabello real, sino a algo más que podríamos llamar cabello' (“cabello virtual”). Así que el hecho de que Neo no tenga cabello real es irrelevante con respecto a si su creencia es verdadera. Neo realmente tiene cabello virtual, por lo tanto tiene razón. Igualmente, cuando un niño en la película le dice a Neo “no hay ninguna cuchara”, su concepto se refiere a una cuchara virtual, y hay realmente una cuchara virtual. De ahí que el niño esté equivocado.

Objeción 4: ¿A qué *clase* de objetos está haciendo referencia un ser encubetado? ¿Qué es cabello virtual, Tucson virtual, y así sucesivamente?

Respuesta: éstas son entidades constituidas por procesos computacionales. Si estoy encubetado, entonces los objetos a que me refiero (cabello, Tucson, y así sucesivamente), están todos hechos de bits. Y si otro ser está encubetado, los objetos a que se refiere (cabello*, Tucson*, y así sucesivamente) igualmente están hechos de bits. Si el ser encubetado está conectado a una simulación en mi computador, entonces los objetos a los que se refiere están constituidos por patrones de bits dentro de mi computadora. Podríamos llamar a estas cosas *objetos virtuales*. Las manos virtuales no son manos (asumiendo que no estoy encubetado), pero ellas existen dentro de la computadora al mismo tiempo. El Tucson virtual no es Tucson, pero existe dentro de la computadora al mismo tiempo.

Objeción 5: usted sólo dijo que las manos virtuales no son manos reales. Esto significa que si estamos en la matrix, ¿no tenemos manos reales?

Respuesta: no. Si *no* estamos en la matrix, pero alguien más está, podemos decir que su término “mano” se refiere a manos virtuales, pero nuestro término no. De ahí que en este caso, nuestras manos no son manos virtuales. Pero si *estamos* en la matrix, entonces nuestro término “mano” se refiere a algo hecho de bits: manos virtuales, o por lo menos a algo que sería considerado como manos virtuales por personas en el mundo contiguo superior. Es decir, si *estamos* en la matrix, las manos reales están hechas de bits. Las cosas parecen bastante diferentes, y nuestras palabras se refieren a cosas diferentes, dependiendo de si nuestra perspectiva es desde dentro o fuera de la matrix.

Esta clase de cambio de perspectiva es común en el pensamiento acerca del escenario de la matrix. Desde la perspectiva de la primera persona, suponemos que *estamos* en una matrix. Aquí las cosas reales en nuestro mundo están hechas de bits, aunque el “mundo contiguo superior” podría no estar hecho de bits. Desde la perspectiva de la tercera persona, suponemos que *algo más* está en una matrix, pero nosotros no. Aquí las cosas reales en nuestro mundo no están hechas de bits, pero el “mundo contiguo inferior” está hecho de bits. En la primera manera de hacer cosas, nuestras palabras se refieren a entidades computacionales. En la segunda manera de hacer cosas, las palabras de los seres encubetados se refieren a entidades computacionales, pero nuestras palabras no.

Objeción 6: realmente ¿cuáles patrones de bits son un objeto virtual dado? Ciertamente será imposible escoger un conjunto preciso.

Respuesta: esta pregunta equivale a preguntarse: realmente ¿cuál parte de las funciones de onda de los quantum es esta silla o es la Universidad de Arizona? Todos estos objetos están, en última instancia, constituidos por unas funciones de onda de quantum subyacentes, pero podría no haber una porción precisa de la función de onda del micro nivel de la que podamos decir que “es” la silla o la universidad. La silla y la universidad existen en un nivel más alto. Igualmente, si estamos encubetados, puede no haber ningún conjunto preciso de bits en los procesos computacionales del micro nivel que sean la silla o la universidad. Éstas existen en un nivel más alto. Y si algunas personas están encubetadas, podría no haber ningún conjunto preciso de bits en la simulación computacional que “sean” los objetos a los que ellas se refieren. Pero así como una silla existe sin ser una parte precisa de las

funciones de onda, una silla virtual puede existir sin ser un conjunto preciso de bits.

Objeción 7: un ser encubetado piensa que realiza sus acciones, y piensa que tiene amigos. ¿Son estas creencias correctas?

Respuesta: uno podría intentar decir que el ser realiza acciones' y que tiene amigos'. Pero por varias razones pienso que no es plausible que palabras como "acción" y "amigo" puedan cambiar sus significados tan fácilmente como palabras de la clase de "Tucson" y "cabello". En cambio, pienso que uno puede decir verídicamente (en nuestro propio lenguaje) que el ser encubetado realiza acciones y tiene amigos. Seguramente, realiza sus acciones en *su* entorno, y su entorno no es nuestro entorno, sino el entorno virtual. Igualmente, sus amigos habitan el entorno virtual (asumiendo que tenemos una matrix multi-cubeta, o que el cómputo es suficiente para la conciencia). Pero el ser encubetado no está equivocado con respecto a esto.

Objeción 8: manteniendo estos puntos técnicos a un lado, seguramente, si estamos en una matrix, ¡el mundo no es como pensamos que es!

Respuesta: niego esto. Incluso si estamos en una matrix, todavía hay personas, juegos de fútbol, y todas estas cosas están ubicadas justo en el espacio-tiempo como pensamos que lo están. Es sólo que el mundo tiene una naturaleza *posterior* que va más allá de nuestra concepción inicial. En particular, las cosas en el mundo son realizadas computacionalmente de una manera que podríamos no haber imaginado originalmente. Pero esto no contradice ninguna de nuestras creencias ordinarias. A lo sumo, contradirá algunas de nuestras creencias metafísicas más abstractas. Pero exactamente lo mismo se aplica para la mecánica cuántica, la teoría de la relatividad, y así sucesivamente.

Si estuviéramos en una matrix, podría ocurrir que no tuviéramos muchas creencias falsas, sino más bien que careciéramos de mucho conocimiento. Por ejemplo, no sabríamos que el universo está realizado computacionalmente. Pero esto es justo lo que debería esperarse. Incluso si no estuviéramos en una matrix, podría perfectamente ocurrir que desconociéramos muchas cosas sobre la naturaleza fundamental de la realidad. No somos criaturas omniscientes, y nuestro conocimiento del mundo es, en el mejor de los casos, parcial. Esta es simplemente la condición de una criatura que vive en un mundo.

VIII OTRAS HIPÓTESIS ESCÉPTICAS

La Hipótesis de la Matrix es un ejemplo de una tradicional hipótesis “escéptica”, pero no es el único ejemplo. Otras hipótesis escépticas no son tan simples como la Hipótesis de la Matrix. Sin embargo, pienso que a muchas de ellas se les aplica una línea similar de razonamiento. En particular, uno puede argumentar que la mayoría de éstas no son hipótesis escépticas universales: esto es, su verdad no socavaría todas nuestras creencias empíricas acerca del mundo físico. En el peor de los casos, la mayoría de ellas son hipótesis escépticas *parciales* que socavan algunas de nuestras creencias empíricas, pero que dejan intactas a muchas otras.

Nueva Hipótesis de la Matrix: fui recientemente creado, junto con todos mis recuerdos, y fui puesto en una matrix recientemente creada.

¿Y qué ocurre si tanto la matrix como yo existimos sólo por un breve periodo de tiempo? Esta hipótesis es una versión computacional de la Hipótesis de la Creación Reciente de Bertrand Russell: el mundo físico fue creado sólo recientemente (con los registros fósiles intactos), y también yo (con mis recuerdos intactos). En esta hipótesis, el mundo externo que percibo existe realmente, y la mayoría de mis creencias acerca de sus actuales estados son plausiblemente verdaderas, pero tengo muchas creencias falsas acerca del pasado. Pienso que lo mismo puede decirse de la Nueva Hipótesis de la Matrix. Uno puede argumentar, siguiendo la línea anterior de razonamiento, que la Nueva Hipótesis de la Matrix es equivalente a la combinación de la Hipótesis Metafísica con la Hipótesis de la Creación Reciente. Esta combinación no es una hipótesis escéptica universal (aunque es una hipótesis escéptica parcial, donde las creencias acerca del pasado se ven comprometidas). Lo mismo se aplica para la Nueva Hipótesis de la Matrix.

Hipótesis de la Matrix Reciente: no he estado encubetado la mayor parte de mi vida, pero fui recientemente conectado a una matrix.

Si recientemente fui puesto en una matrix sin darme cuenta, parece que muchas de mis creencias acerca de mi entorno son falsas. Digamos simplemente que justo ayer alguien me introdujo en una simulación, en la que vuelo a Las Vegas y juego en un casino. Entonces, podría creer que ahora estoy en Las Vegas, y que estoy en un casino, pero estas creencias son falsas: estoy realmente en un laboratorio en Tucson.

Este resultado es completamente diferente de la matrix de largo plazo. La diferencia está en el hecho de que mi concepción de la realidad externa

se adapta a la realidad en la que he vivido la mayoría de mi vida. Si he estado encubetado toda mi vida, mi concepción está adaptada a la realidad computacional. Pero si estoy encubetado apenas desde ayer, mi concepción está adaptada a la realidad externa. Cuando pienso que estoy en Las Vegas, estoy pensando que estoy en Las Vegas externa, y este pensamiento es falso.

Sin embargo, esto no socava todas mis creencias sobre el mundo externo. Creo que nací en Sydney, que hay agua en los océanos, y así sucesivamente, y todas estas creencias son correctas. Son sólo mis creencias recientemente obtenidas, las cuales provienen de la percepción del ambiente simulado, las que son falsas. Así, esta es sólo una hipótesis escéptica parcial: su posibilidad plantea dudas sobre un subconjunto de nuestras creencias empíricas, pero no plantea ninguna duda sobre todas ellas.

De manera interesante, la Hipótesis de la Matrix Reciente y la Nueva Hipótesis de la Matrix dan resultados opuestos, a pesar de su similar naturaleza: la Hipótesis de la Matrix Reciente produce creencias verdaderas acerca del pasado, pero falsas acerca del presente, mientras que la Nueva Hipótesis de la Matrix produce creencias falsas sobre el pasado y creencias verdaderas sobre el presente. Las diferencias dependen del hecho de que en la Hipótesis de la Matrix Reciente yo realmente tengo una existencia pasada acerca de la cual son mis creencias, y que la realidad pasada ha desempeñado un papel en la determinación de los contenidos de mis pensamientos que no tiene paralelo alguno bajo la Nueva Hipótesis de la Matrix.

Hipótesis de la Matrix Local: estoy conectado a una simulación computarizada de un entorno local fijo en un mundo.

En una de las versiones de esta idea, una simulación computarizada simula un entorno pequeño fijo en un mundo, y los sujetos en la simulación encuentran alguna clase de barrera cuando intentan dejar esa área. Por ejemplo, en la película *The Thirteenth Floor*, sólo California es simulada, y cuando el sujeto intenta conducir a Nevada, en la carretera dice “Cerrado por Reparación” (¡con tenues montañas verdes electrónicas en la distancia!). Por supuesto, esta no es la mejor manera de crear una matrix, cuando es probable que el sujeto esté descubriendo los límites de su mundo.

Esta hipótesis es análoga a una Hipótesis de la Creación Local, en que los creadores sólo crean una parte local del mundo físico. Bajo esta

hipótesis, tendremos creencias verdaderas sobre los objetos cercanos, pero creencias falsas sobre las situaciones lejanas. Por la clase usual de razonamiento, la Hipótesis de la Matrix Local puede ser vista como una combinación de la Hipótesis Metafísica con la Hipótesis de la Creación Local. Así, diríamos lo mismo sobre esto.

La Hipótesis de la Matrix Localmente Extensible: estoy conectado a una simulación computarizada de un entorno local fijo en un mundo, extendido cuando es necesario dependiendo de los movimientos del sujeto.

Esta hipótesis evita las dificultades obvias de una matrix local fija. Aquí los creadores simulan un entorno local y lo extienden cuando es necesario. Por ejemplo, podrían estar concentrándose ahora en simular una habitación en mi casa en Tucson. Si entro en otra habitación, o si vuelo a otra ciudad, ellos las simularán. Por su puesto, necesitan asegurarse que cuando vaya a estos lugares, mis recuerdos y creencias estarán unidos razonablemente bien, con posibilidades para evolucionar en tiempos posteriores. Lo mismo se aplica cuando encuentro personas familiares, o personas sobre las que sólo he oído. Presumiblemente, los simuladores mantienen un banco de datos de la información acerca del mundo que ha sido estable hasta ahora, actualizando esta información cada vez que sea necesario cuando el tiempo avance, y haciendo nuevos detalles cuando lo necesiten.

Esta clase de simulación es completamente diferente de una simulación en una matrix ordinaria. En una matrix, el mundo entero es simulado de inmediato. Hay altos costos para empezar, pero una vez la simulación es puesta en marcha, cuidará de sí misma. Por contraste, la matrix local extensible involucra una simulación "justo-a-tiempo". Esto tiene costos mucho más bajos para empezar, pero requiere mucho más trabajo y creatividad en la medida en que la simulación se desarrolla.

Esta hipótesis es análoga a una Hipótesis de la Creación Localmente Extensible acerca de la realidad ordinaria, bajo la que los creadores crean sólo un entorno físico local, y lo extienden cuando es necesario. Aquí, la realidad externa existe y muchas creencias locales son verdaderas, pero de nuevo las creencias acerca de las cosas lejanas son falsas. Si combinamos esta hipótesis con la Hipótesis Metafísica, el resultado es la Hipótesis de la Matrix Localmente Extensible. Así, si estamos en una matrix local extensible, la realidad externa aún existe, pero no hay tanto de ella como pensamos. Por su puesto, si viajo en la dirección correcta, ¡más de esta realidad comienza a existir!

Esta situación recuerda la de *The Truman Show*. Truman vive en un entorno artificial compuesto de actores y accesorios, que se comporta adecuadamente cuando él está alrededor, pero que puede ser completamente diferente cuando él está ausente. Truman tiene muchas creencias verdaderas acerca de su entorno actual: realmente hay mesas y sillas enfrente de él, y así sucesivamente. Pero está profundamente equivocado acerca de las cosas fuera de su entorno actual, y sobre las lejanas.

Es común pensar que mientras *The Truman Show* propone un inquietante escenario escéptico, *The Matrix* es mucho peor. Pero si estoy en lo cierto, las cosas se invierten. Si estoy en una matrix, entonces la mayoría de mis creencias acerca del mundo externo son verdaderas. Si estoy en algo como *The Truman Show*, entonces un gran número de mis creencias son falsas. Pensando sobre esto, me parece que esta es la conclusión correcta. Si fuéramos a descubrir que estábamos (y siempre hemos estado) en una matrix, esto sería sorprendente, pero podríamos rápidamente acostumbrarnos a ella. Si descubrimos que estábamos (y siempre hemos estado) en el show de Truman, podríamos volvernos locos.

La Hipótesis de la Matrix Macroscópica: estoy conectado a una simulación computarizada de procesos físicos macroscópicos sin detalles microfísicos.

Uno puede imaginar que por facilidad de la simulación, los fabricantes de una matrix no pueden simular la física de bajo nivel. En cambio, sólo pueden representar objetos macroscópicos en el mundo y sus propiedades: e.g. que hay una mesa con tal y tal forma, posición y color, con un libro en cima y con ciertas propiedades, y así sucesivamente. Ellos pueden necesitar hacer algunos esfuerzos para asegurarse que estos objetos se comporten de una manera físicamente razonable, y tendrán que hacer provisiones especiales para manejar medidas microfísicas, pero uno puede imaginar que por lo menos una simulación razonable puede ser creada de esta manera.

Pienso que esta hipótesis es análoga a una Hipótesis del Mundo Macroscópico: no hay procesos microfísicos, y en cambio los objetos físicos macroscópicos existen como objetos fundamentales en el mundo, con propiedades de forma, color, posición, y así sucesivamente. Esta es una manera coherente como nuestro mundo podría ser, y no es una hipótesis escéptica global, aunque puede llevar a creencias científicas falsas sobre los niveles más bajos de la realidad. La Hipótesis

Macroscópica de la Matrix puede ser vista como una combinación de esta hipótesis con una versión de la Hipótesis Metafísica. Como tal, esta no es otra hipótesis escéptica global.

Uno puede también combinar las hipótesis que hemos mencionado antes de diversas maneras, produciendo hipótesis tales como una Hipótesis de la Matrix Local Macroscópica Nueva. Por las razones usuales, todas éstas pueden ser vistas como análogas de hipótesis correspondientes acerca del mundo físico. De esta manera, todas ellas son compatibles con la existencia de la realidad física, y ninguna es una hipótesis escéptica global.

La Hipótesis de Dios: la realidad física está representada en la mente de Dios, y nuestros propios pensamientos y percepciones dependen de la mente de Dios.

Una hipótesis como ésta fue propuesta por George Berkeley como una concepción sobre cómo podría ser realmente nuestro mundo. Berkeley la propuso como una clase de hipótesis metafísica sobre la naturaleza de la realidad. Muchos otros filósofos han diferido de Berkeley con respecto a esta clase de hipótesis escéptica. Si tengo razón, Berkeley está más cerca de la verdad. La Hipótesis de Dios puede ser vista como una versión de la Hipótesis de la Matrix, en que la simulación del mundo es implementada por la mente de Dios. Si esto es correcto, podemos decir que los procesos físicos realmente existen: sólo que en un nivel más fundamental, están constituidos por procesos en la mente de Dios.

La Hipótesis del Genio Maligno: tengo una mente separada del cuerpo, y un genio maligno me estimula sensorialmente para darme la apariencia de un mundo externo.

Esta es la clásica hipótesis escéptica de René Descartes. ¿Qué podemos decir sobre ésta? Esto depende sólo de cómo trabaja el genio maligno. Si el genio maligno simula un mundo entero en su cabeza con el fin de determinar qué estímulos puedo recibir, entonces tenemos una versión de La Hipótesis de Dios. Aquí podemos decir que la realidad física existe y está constituida por procesos dentro del genio. Si el genio maligno está simulando sólo una pequeña parte del mundo físico, solo lo suficiente para darme estímulos razonablemente consistentes, entonces tenemos algo análogo a la Hipótesis de la Matrix Local (en cualquiera de sus dos versiones, la fija o la flexible). Aquí podemos decir que sólo una parte

local de la realidad externa existe. Si el genio maligno no se molesta en simular el nivel microscópico, sino sólo el nivel macroscópico, entonces tenemos algo análogo a la Hipótesis de la Matrix Macroscópica. En este punto, podemos decir que los objetos macroscópicos externos locales existen, pero nuestras creencias acerca de su naturaleza microfísica son incorrectas.

La hipótesis del genio maligno es tomada a menudo como una hipótesis escéptica global. Pero si el razonamiento anterior es correcto, esto es falso. Incluso si la Hipótesis del Genio Maligno es correcta, algo de la realidad externa que aparentemente percibimos existe realmente, aunque podamos tener algunas creencias falsas acerca de ella, dependiendo de los detalles. Es sólo que esta realidad externa tiene una naturaleza subyacente que es completamente diferente de lo que podemos haber pensado.

La Hipótesis del Sueño: ahora y siempre he estado soñando.

Descartes planteó la pregunta: ¿cómo sabe usted que en este momento no está soñando? Morpheus plantea una pregunta similar: ¿Has tenido alguna vez un sueño, Neo, del que estuvieras completamente seguro que era real? ¿Y qué si fueras incapaz de despertarte del sueño? ¿Cómo podrías conocer la diferencia entre el sueño y el mundo real? La hipótesis de que estoy *actualmente* soñando es análoga a una versión de la Hipótesis de la Matrix Reciente. No la puedo descartar de manera concluyente, y si es correcta, entonces muchas de mis creencias acerca de mi entorno actual son falsas. Pero, presumiblemente todavía tengo muchas creencias verdaderas acerca del mundo externo, fijadas en el pasado.

¿Y qué si he estado soñando siempre? Esto es, ¿qué pasaría si todos mis estímulos perceptuales aparentes han sido generados por mi propio sistema cognitivo sin que yo me dé cuenta? Pienso que este caso es análogo a la Hipótesis del Genio Maligno: ¡sólo que el papel del “genio maligno” lo desempeña ahora una parte de mi propio sistema cognitivo! Si mi sistema generador de sueños simula todo el espacio-tiempo, tenemos algo como la Hipótesis de la Matrix original. Si modela sólo mi entorno local, o sólo algunos procesos macroscópicos, tenemos algo análogo de las versiones más locales de la Hipótesis del Genio Maligno. En cualquiera de estos casos, podemos decir que los objetos que actualmente percibimos realmente existen (aunque los objetos más lejanos puede que no). Sólo que algunos de ellos están constituidos por mis propios procesos cognitivos.

La Hipótesis del Caos: no recibo estímulos de ninguna parte del mundo. En cambio, tengo experiencias aleatorias incausadas. Gracias a una enorme coincidencia, ellas son de la clase de experiencias regulares y estructuradas que me son familiares.

La Hipótesis del Caos es extraordinariamente improbable, mucho más improbable que cualquiera de las hipótesis consideradas hasta ahora. Aún así, podría, en principio, ser verdadera a pesar de su minúscula probabilidad. Si estoy encubetado caóticamente, ¿se dan procesos físicos en el mundo externo? Pienso que debemos decir que no se dan. Mis experiencias de los objetos externos no son causadas por nada, y el grupo de experiencias asociadas con mi concepción de un objeto dado podría no tener ninguna fuente común. Ciertamente, mis experiencias no están causadas en lo absoluto por ninguna realidad externa a ellas. Esta sí es una hipótesis escéptica genuina: si se la acepta, tendrían que rechazarse la mayoría de nuestras creencias acerca del mundo externo.

Hasta ahora, el único caso claro de una hipótesis escéptica global es la Hipótesis del Caos. A diferencia de las hipótesis anteriores, la aceptación de esta hipótesis socavaría todas nuestras creencias sustantivas sobre el mundo externo. ¿De dónde proviene la diferencia?

Posiblemente, lo que es crucial es que en la Hipótesis del Caos no hay ninguna explicación causal de nuestras experiencias, y no hay explicación para las regularidades de nuestra experiencia. En todos los casos previos, hay alguna explicación para estas regularidades, aunque quizás no la explicación que esperamos. Uno podría sugerir que en la medida en que una hipótesis involucre *alguna* explicación razonable para las regularidades de nuestra experiencia, entonces no será una hipótesis escéptica global.

Si es así, entonces si estamos consintiendo la suposición de que hay algunas explicaciones para las regularidades de nuestras experiencias, no hay problema en decir que algunas de nuestras creencias acerca del mundo externo son correctas. Esto no es mucho, ¡pero es algo!

IX NOTAS FILOSÓFICAS

Este artículo fue escrito con el propósito de ser accesible para una audiencia amplia, por eso omite deliberadamente los detalles filosóficos

técnicos, las referencias a la literatura, y así sucesivamente. En estas notas intentaré remediar esta omisión. Los lectores sin formación en filosofía probablemente deben pasar por alto esta sección.

Nota 1: Hilary Putnam (1981) argumentó que la hipótesis de que soy (y siempre he sido) un cerebro en una cubeta puede ser desechada a priori. En efecto, esto es porque mi palabra “cerebro” se refiere a objetos en mi mundo percibido, y no puede referirse a objetos en un mundo “externo” en el que tendría que existir la cubeta. Para que mi hipótesis “soy un cerebro en una cubeta” sea verdadera, yo tendría que ser un cerebro de la clase que exista en el mundo percibido, pero este no puede ser el caso. Por lo tanto, la hipótesis debe ser falsa.

Una analogía: puede rechazarse razonablemente la hipótesis de que estoy en la Matrix (con M mayúscula). Mi término “la Matrix” se refiere a un sistema específico que he visto en una película en mi mundo percibido. Yo no podría estar en ese mismo sistema, ya que el sistema existe dentro del mundo que percibo. Entonces, mi hipótesis “estoy en la Matrix” debe ser falsa.

Esta conclusión acerca de la Matrix parece razonable, pero hay una respuesta natural. Quizás este argumento niega la hipótesis de que estoy en la Matrix, pero no puedo negar la hipótesis de que estoy en una matrix, donde una matrix es un término genérico para una simulación computarizada de un mundo. El término “Matrix” podría referirse únicamente al sistema específico en la película, pero el término genérico “matrix” no.

De manera similar, puedo descartar razonablemente la hipótesis de que soy un cerebro en una cubeta (si “cerebro” se refiere únicamente a una clase específica de sistema biológico en mi mundo percibido). Pero no puedo negar la hipótesis de que estoy encubetado, si esto simplemente significa que tengo un sistema cognitivo que recibe estímulos de, y envía respuestas a, una simulación computarizada de un mundo. El término “encubetado” (y los términos usados en su definición) son términos genéricos, no anclados a sistemas específicos en la realidad percibida. Al usar este lenguaje ligeramente distinto, podemos reformular la hipótesis escéptica de tal manera que resulte invulnerable para el argumento de Putnam.

Más técnicamente: el argumento de Putnam puede funcionar para “cerebro” y “Matrix” porque el uno es un término de género natural y el otro es un nombre propio. Estos términos están sujetos a los experimentos mentales de la “Tierra Gemela” (Putnam 1975), donde los duplicados pueden usar términos correspondientes con distinta referencia. En la Tierra, el término “agua” de Oscar se refiere a H_2O , pero en la Tierra Gemela no (que contiene la sustancia superficialmente idéntica XYZ en sus océanos y lagos). El término “agua” del gemelo de Oscar se refiere a XYZ. Igualmente, quizás mi término “cerebro” se refiere a cerebros biológicos, mientras que el término “cerebro” para un ser encubetado se refiere a cerebros virtuales. En este caso, cuando un ser encubetado dice “soy un cerebro en una cubeta”, no se está refiriendo a su cerebro biológico, y su afirmación es falsa.

Pero no todos los términos están sujetos a los experimentos mentales de la Tierra Gemela. En particular, los términos *semánticamente neutrales* no lo están (por lo menos cuando son usados sin deferencias semánticas): tales términos plausiblemente incluyen “filósofo”, “amigo”, y muchos otros. Otros términos semejantes incluyen “matrix” y “encubetado”, como están definidos en este artículo. Si trabajamos con hipótesis tales como “estoy en una matrix” y “estoy encubetado”, en lugar de “estoy en la Matrix” o “soy un cerebro en una cubeta”, entonces el argumento de Putnam no se aplica. Incluso si un cerebro en una cubeta no puede realmente pensar “soy un cerebro en una cubeta”, puede realmente pensar “estoy encubetado”. Pienso que la línea de razonamiento de Putnam es en última instancia una pista falsa.

Nota 2: a pesar de este desacuerdo, la conclusión de este artículo está estrechamente relacionada con otra sugerencia de Putnam. Esta es la sugerencia de que un cerebro en una cubeta puede tener creencias verdaderas, porque podría referirse a procesos químicos o a procesos dentro de una computadora. Sin embargo, llego a esta conclusión por una ruta completamente diferente. Putnam argumenta por una apelación a la teoría causal de la referencia: los pensamientos se refieren a aquello con lo que están causalmente conectados, y los pensamientos de un ser encubetado están causalmente conectados a procesos en una computadora. Este argumento es claramente inconclusivo, en la medida en que la teoría causal de la referencia no está todavía restringida. Afirmar que una conexión causal se requiere para la referencia no es decir qué clase de conexión causal es suficiente. Hay muchos casos (como el del “flogisto”) donde los términos fracasan al referirse a pesar de las ricas conexiones causales. Intuitivamente, es natural pensar que

el cerebro en una cubeta es un caso como éste. Por esto una apelación a la teoría causal de la referencia no parece ayudar.

El argumento que he ofrecido no presupone nada acerca de la teoría de la referencia. Más bien, procede directamente mediante la consideración de hipótesis de primer orden acerca del mundo, las conexiones entre éstas, y lo que diríamos si fueran verdaderas. Al responder a las objeciones, he hecho algunas observaciones sobre la referencia, las cuales son ampliamente compatibles con una teoría causal de la referencia. Pero lo más importante es que estas afirmaciones son, en gran medida, consecuencias del argumento del primer orden, antes que suposiciones de éste. En general, creo que las afirmaciones de la teoría de la referencia están respaldadas en los juicios de primer orden acerca de los casos, antes que al contrario.

Nota 3: uso en este artículo “hipótesis escéptica” en un cierto sentido técnico. Una hipótesis escéptica (relativa a una creencia de que P) es una hipótesis tal que (i) no podemos descartarla con certeza, (ii) si la aceptáramos deberíamos rechazar la creencia de que P. Una hipótesis escéptica con respecto a una clase de creencias es tal que es una hipótesis escéptica con respecto a la mayoría de, o a todas las creencias de la clase. Una hipótesis escéptica global es una hipótesis escéptica con respecto a todas nuestras creencias empíricas.

La existencia de una hipótesis escéptica (con respecto a una creencia) proyecta dudas sobre la creencia relevante, en el siguiente sentido. Debido a que no podemos rechazar la hipótesis con certeza, y debido a que la hipótesis implica la negación de estas creencias, parece (dado un principio de cierre plausible sobre la certeza) que no podemos conocer estas cosas con certeza. Si también ocurre que no *sabemos* que la hipótesis escéptica es falsa (como creo que ocurre con la mayoría de hipótesis en este artículo), entonces se sigue de un principio análogo de cierre que las creencias en la clase relevante no constituyen conocimiento.

Algunos usan “hipótesis escéptica” en un sentido más general, para aplicarlo a cualquier hipótesis tal que si es verdadera, no conozco que P. (una hipótesis bajo la que tengo accidentalmente creencias verdaderas es una hipótesis escéptica en este sentido, pero no en el anterior). No he argumentado aquí que la Hipótesis de la Matrix no es una hipótesis escéptica en este sentido. He argumentado que si la hipótesis es verdadera, nuestras creencias son verdaderas, pero no he argumentado que si es verdadera, nuestras creencias constituyen conocimiento. No

obstante, estoy inclinado a pensar que si tenemos conocimiento en un mundo ordinario no-matrix, podríamos también tener conocimiento en una matrix.

Nota 4: ¿cuál es la clase relevante de creencias? Por supuesto, hay algunas creencias que incluso una hipótesis escéptica de un mundo no externo no podría socavar: la creencia de que existo, o la creencia de que $2+2=4$, o la creencia de que no hay unicornios. Debido a esto, es mejor restringir la atención a creencias que (i) sean sobre el mundo externo, (ii) no sean justificables a priori, y (iii) hagan una afirmación positiva acerca del mundo (ellas no serían verdaderas en un mundo vacío). Para los propósitos de este artículo, podemos pensar en estas creencias como nuestras “creencias empíricas”. Las afirmaciones acerca de las hipótesis escépticas que socavan creencias pueden generalmente ser entendidas como restringidas a las creencias en esta clase.

Nota 5: sobre la Hipótesis Computacional: es coherente suponer que hay un nivel computacional debajo de lo físico, pero no es claro si es coherente suponer que este nivel es fundamental. Si fuera fundamental, entonces tendríamos un mundo de “puros bits”. Un mundo así sería un mundo de diferencias puras: hay dos estados básicos que difieren uno del otro, sin que esta diferencia sea una diferencia en alguna naturaleza más profunda. Que esto sea coherente o no, depende de si uno cree que todas las diferencias deben estar basadas en alguna naturaleza intrínseca básica, de si uno piensa que todas las disposiciones deben tener unas bases categóricas, y así sucesivamente. Para los propósitos de este artículo, sin embargo, el problema puede dejarse a un lado. Bajo la Hipótesis de la Matrix, el propio cómputo está *implementado* por procesos en el mundo del creador. En este sentido, habrá un nivel más básico de propiedades intrínsecas que sirve como la base de las diferencias entre los bits.

Nota 6: sobre la Hipótesis Mente-Cuerpo: es interesante anotar que la Hipótesis de la Matrix muestra una manera concreta en que el dualismo cartesiano de substancias podría ser verdadero. A veces se sostiene que la idea de que procesos físicos interactúen con una mente no física no es sólo implausible sino incoherente. La Hipótesis de la Matrix sugiere abiertamente que esto es erróneo. Bajo esta hipótesis, nuestro sistema cognitivo involucra procesos completamente distintos de los procesos en el mundo físico, pero hay una clara historia causal acerca de cómo interactúan ambas clases de procesos.

Surgen algunas preguntas. Por ejemplo, si el sistema cognitivo encubetado produce respuestas motoras del cuerpo, ¿qué papel desempeña el cerebro simulado? Quizás uno podría no contar con él, pero esto puede causar toda una clase de resultados extraños, no menos cuando los doctores en la matrix abren el cráneo. Es más natural pensar que el cerebro encubetado y el cerebro simulado estarán siempre en estados isomórficos, recibiendo los mismos estímulos y produciendo las mismas respuestas. Si los dos sistemas comienzan en estados isomórficos y siempre reciben los mismos estímulos, entonces (dejando a un lado el indeterminismo) siempre permanecerán en estados isomórficos. Como suplemento, ¿esto puede explicar por qué la muerte en la Matrix lleva a la muerte en el mundo exterior!

¿Cuáles de estos realmente controlan el cuerpo? Esto depende de cómo las cosas se establezcan. Las cosas pueden ser establecidas para que las respuestas del sistema encubetado no se retroalimenten a la simulación; en este caso una versión del epifenomenalismo será verdadera. Las cosas pueden ser establecidas para que los impulsos motores en el cuerpo simulado dependan de las respuestas del sistema encubetado, ignorando las respuestas del cerebro simulado; en este caso una versión del interaccionismo será verdadera. De manera interesante, ¿esta última puede ser una versión del interaccionismo que es compatible con el cierre causal de lo físico! Una tercera posibilidad es que los mecanismos tienen ambas clases de respuestas (¿quizás promediando las dos?). Esto podría conducir a una suerte de redundancia en la causación. Quizás los controladores de la matrix podrían, incluso algunas veces, alternar entre las dos. En cualquiera de estos casos, con tal de que los dos sistemas permanezcan en estados isomórficos, los resultados conductuales serán los mismos.

Uno podría preocuparse de si existen aquí dos mentes conscientes, en un modo que recordara la historia de Daniel Dennett ¿“Dónde estoy”? Esto depende de si el cómputo en la matrix es suficiente para soportar una mente. Si el anti-computacionalismo acerca de la mente (como el de John Searle) es correcto, habrá simplemente una mente. Si el computacionalismo acerca de la mente es correcto, bien puede haber dos mentes sincronizadas (lo cual, entonces, plantea la pregunta: si estoy en la matrix, ¿cuál de las mentes es la mía?). El punto de vista de que hay una mente es más cercano a la concepción ordinaria de la realidad, pero el punto de vista de que hay dos mentes no está completamente fuera de lugar.

Una ventaja adicional del punto de vista computacionalista es que nos permite adoptar la hipótesis de que estamos en una simulación computarizada *sin* un sistema cognitivo separado vinculado a ella. En cambio, los creadores simplemente ejecutan la simulación, incluyendo una simulación de cerebros y mentes que emergen adentro. Esto es presumiblemente mucho más fácil para los creadores, ya que elimina cualquier problema con respecto a la creación y mantenimiento de los sistemas cognitivos encubetados. Debido a esto, es muy plausible que haya muchas simulaciones de esta clase en el futuro, mientras que es incierto que haya muchas más de las embarazosas simulaciones estilo Matrix. (Por tal motivo, el argumento de Bostrom de que muy bien podría ser que estuviéramos en una simulación, se aplica más directamente a esta clase de simulación que a una estilo Matrix). La hipótesis de que estamos en esta clase de simulación computarizada corresponde a una versión debilitada de la Hipótesis Metafísica, en la que la Hipótesis Mente-Cuerpo es innecesaria. Como antes, esta no es una hipótesis escéptica: si estamos en semejante simulación (y si el computacionalismo sobre la mente es verdadero), entonces la mayoría de nuestras creencias sobre el mundo externo todavía son correctas.

Hay también otras posibilidades. Una que es intrigante (discutida en Chalmers 1990) es sugerida por el trabajo contemporáneo en vida artificial que involucra entornos simulados relativamente simples, y reglas complejas por las que las criaturas simuladas interactúan con estos entornos. Aquí los algoritmos responsables por los procesos “mentales” de las criaturas son completamente distintos de aquellos que gobiernan la “física” del entorno. En esta clase de simulación, las criaturas presumiblemente nunca encontrarán soporte para sus procesos cognitivos en su mundo percibido. Si estas criaturas se hicieran científicos, serían dualistas cartesianos, sosteniendo (¡correctamente!) que sus procesos cognitivos yacen fuera de su mundo físico. Parece que esto es otra forma coherente en que el dualismo cartesiano puede resultar verdadero.

Nota 7: he argumentado que la Hipótesis de la Matrix implica la Hipótesis Metafísica y viceversa. Aquí, “implica” es una relación epistémica: si uno acepta la primera, debería aceptar la segunda. No afirmo que la Hipótesis de la Matrix *entrañe* la Hipótesis Metafísica, en el sentido de que en cualquier mundo contrafáctico, en el cual sea verdadera la Hipótesis de la Matrix, también lo sea la Hipótesis Metafísica. Esta afirmación parece falsa. Por ejemplo, hay mundos contrafácticos en los

que el espacio-tiempo físico no es creado por nadie (de esta manera, la Hipótesis Metafísica es falsa), en los cuales puedo estar conectado a una simulación computarizada diseñada artificialmente, ubicada dentro del espacio-tiempo físico (de tal modo que la Hipótesis de la Matrix es verdadera). Y si la física no es computacional en el mundo real, entonces la física en este mundo no es computacional tampoco. Uno puede afirmar que las dos hipótesis son equivalentes *a priori*, pero no necesariamente equivalentes.

(Por su puesto, el término "física" como es usado por mi yo encubetado en el mundo contrafáctico se referirá a algo que es a su vez computacional y creado. Pero "física" cuando es usado por mi yo encubetado actual, se refiere a la física exterior no computacional de ese mundo, no a los procesos computacionales).

La diferencia surge de dos formas diferentes de considerar la Hipótesis de la Matrix: como una hipótesis sobre lo que podría ocurrir realmente, o como una hipótesis sobre lo que no es pero podría haber sido. La primera hipótesis está reflejada en condicionales indicativos: si estoy actualmente en una matrix, entonces tengo manos, los átomos están hechos de bits, y la Hipótesis Metafísica es verdadera. La segunda versión está reflejada en los condicionales subjuntivos: si hubiera estado en una matrix, podría no haber tenido manos, y los átomos no estarían hechos de bits, y la Hipótesis Metafísica no sería verdadera.

Esto es análogo a las diferentes formas de pensar sobre la Tierra Gemela de Putnam, comunes a las discusiones de la semántica bidimensional. Si estoy actualmente en el mundo XYZ, entonces XYZ es agua; pero si hubiera estado en el mundo XYZ, XYZ no habría sido agua (el agua seguiría siendo H₂O). En la primera interpretación, consideramos el mundo de la Tierra Gemela *como real*. En la segunda interpretación, lo consideramos *como contrafáctico*. Podemos decir que el mundo de la Tierra Gemela *verifica* el enunciado "el agua es XYZ", pero que *satisface* el enunciado "el agua no es XYZ", donde la verificación y la satisfacción se corresponden con como real y como contrafáctico.

Similarmente, podemos decir que un mundo matrix verifica la Hipótesis Metafísica, pero no la satisface. La razón es que la Hipótesis Metafísica hace afirmaciones sobre la física y el mundo físico. Y lo que cuenta como "física" será diferente dependiendo de si el mundo matrix es considerado como real o como contrafáctico. Si estoy en una matrix, la física es

computacional. Pero si *hubiera estado* en una matrix, la física podría no haber sido computacional (la matrix podría haber sido computacional, pero la computadora y mi cerebro habrían estado hechos de una física independiente del cómputo). De esta manera, las afirmaciones sobre la física y los procesos físicos en el mundo matrix son análogas a las afirmaciones sobre el “agua” en el mundo de la Tierra Gemela.

Nota 8: las respuestas a las primeras pocas objeciones en la sección 7 son claramente análogas a la descripción causal de la referencia. He afirmado que la verdad de los pensamientos de un ser encubetado no depende de su entorno inmediato, sino de qué es lo que está causalmente conectado con qué: esto es, en los procesos computacionales a los que está conectado. Como he anotado anteriormente, no necesito asumir la teoría causal de la referencia para obtener esta conclusión. Pero, en cambio, llego allí a través de un argumento de primer orden. Pero, cuando la conclusión es alcanzada, hay muchos puntos de contacto interesantes.

Por ejemplo, la idea de que mi término “cabello” se refiere al cabello, mientras el término de mi contraparte encubetada se refiere al cabello virtual, tiene una estructura familiar. Esto es estructuralmente análogo a un caso de Tierra Gemela, en que Oscar (en la Tierra) se refiere al agua (H_2O), mientras su contraparte Oscar Gemelo (en la Tierra Gemela) se refiere al agua gemela (XYZ). En ambos casos, los términos se refieren a con qué están causalmente conectados. Estos términos de género natural funcionan escogiendo una cierta clase en el entorno del sujeto, y la naturaleza precisa de que esta clase depende de la naturaleza del ambiente. Algo similar se aplica a los nombres para las entidades específicas, tales como “Tucson”.

El comportamiento de estos términos puede ser modelado usando el marco de la semántica bidimensional. Como dijimos antes, cuando consideramos un mundo de la Tierra Gemela como real, éste verifica el enunciado “el agua es XYZ”, y cuando lo consideramos como contrafáctico, satisface el enunciado “el agua no es XYZ”. Igualmente, cuando consideramos un mundo matrix como real, éste verifica el enunciado “el cabello está hecho de bits”, y cuando lo consideramos como contrafáctico, satisface el enunciado “el cabello no está hecho de bits”.

La diferencia entre considerarlo como real y contrafáctico proporciona un cambio de perspectiva como aquel que se dio como respuesta en la objeción 5. Si el mundo matrix es considerado como meramente

contrafáctico, debemos decir que los seres en la matrix no tienen cabello (sólo tienen cabello virtual). Pero si el mundo matrix es considerado como real (esto es, si hipotéticamente aceptamos que estamos en una matrix), debemos decir que los seres en la matrix tienen cabello, y que el cabello es en sí mismo una clase de cabello virtual.

La analogía de la tierra gemela puede sugerir que los significados de nuestros términos como “cabello”, y los contenidos de nuestros pensamientos correspondientes dependen de nuestro entorno. Pero la explicación bidimensional también sugiere que hay un aspecto interno del contenido que es compartido entre los gemelos, y que no depende del entorno. La *intensión primaria* de una oración en un mundo es verdadera si el mundo verifica la oración, mientras su *intensión secundaria* es verdadera en un mundo si el mundo satisface la oración. Entonces las oraciones de Oscar y el gemelo de Oscar “el agua es húmeda” tienen diferentes intensiones secundarias (cuando el H₂O es húmeda o cuando lo XYZ es respectivamente húmeda), pero tienen las mismas intensiones primarias (aproximadamente, verdadera en los mundos donde la cosa con aspecto acuoso es húmeda). Igualmente, “tengo cabello”, como es usado por mí y mi contraparte encubetada tiene diferentes intensiones secundarias (aproximadamente, verdadera en mundos donde tengamos respectivamente cabello biológico o cabello computacional), pero tienen la misma intención primaria (aproximadamente, verdadera en mundos donde tengamos cosas con aspecto de cabello). Las intensiones primarias de nuestro pensamiento y nuestro lenguaje representan una considerable dimensión de contenido compartido.

Nota 9: ¿por qué la respuesta es diferente a la objeción 7, sobre “acción” y “amigo”? Notamos antes (nota 1) que no todos los términos funcionan como “agua” y “cabello”. Hay muchos términos *semánticamente neutrales* que no están sujetos a los experimentos mentales de la Tierra Gemela: cualquiera de los dos gemelos usará estos términos con el mismo significado en diferentes entornos (por lo menos si están usando los términos sin deferencias semánticas). Estos términos podrían incluir “y”, “amigo”, “filósofo”, “acción”, “experiencia”, y “encubetado”. Mientras los términos “mano” o “cabello” o “Tucson” de un ser encubetado pueden significar algo diferente de nuestros términos correspondientes. Los términos “amigo” o “filósofo” o “acción” de un ser encubetado posiblemente significarán lo mismo que los nuestros.

Se sigue que si estamos interesados en la creencia “tengo amigos”, o “realizo acciones” de un ser encubetado, no podemos usar la respuesta de la Tierra Gemela. Estas creencias serán verdaderas si y sólo si el ser encubetado tiene amigos y realiza acciones. Afortunadamente, parece completamente razonable afirmar que el ser encubetado tiene amigos (en su entorno, no en el nuestro), y que realiza acciones (en su entorno, no en el nuestro). Lo mismo para otros términos semánticamente neutrales: esta respuesta es razonable precisamente por esta clase de expresiones.

Nota 10: ¿cuál es la ontología de los objetos virtuales? Esta es una pregunta difícil, pero no más difícil que la pregunta por la ontología de los objetos macroscópicos ordinarios en un mundo de mecánica cuántica. La respuesta a la objeción 6 sugiere que en ambos casos deberíamos rechazar afirmaciones de identidad caso a caso entre los niveles micro y macroscópico. Las mesas no son idénticas a ningún objeto caracterizado únicamente en términos de mecánica cuántica; igualmente, las mesas virtuales no son idénticas a ningún objeto caracterizado únicamente en términos de bits. No obstante, los hechos acerca de las mesas sobrevienen a los hechos mecánico- cuánticos, y los hechos acerca de las mesas virtuales sobrevienen a los hechos computacionales. Parece razonable decir que las mesas están constituidas por procesos cuánticos, y que las mesas virtuales están constituidas por procesos computacionales. Un detalle mayor, en cualquiera de los dos casos, depende de delicadas cuestiones metafísicas.

Reflexionando sobre el caso de la tercera persona, en el que estamos mirando a un cerebro en una cubeta en nuestro mundo, uno podría objetar que los objetos virtuales no existen realmente: no hay *objetos* reales en algún lugar que correspondan a las mesas dentro de un computador. Si uno dice esto, sin embargo, puede verse forzado por la similitud con el punto de vista de que las mesas no existen verdaderamente en nuestro mundo mecánico-cuántico. Si uno adopta una ontología restringida de objetos en un caso, debería hacerlo en el otro; si uno adopta una ontología liberal en un caso, debería hacerlo en el otro. La única forma razonable de tratar los casos de un modo diferente es adoptar una clase de contextualismo acerca de qué cuenta como un “objeto” (o sobre lo que cae dentro del dominio de un cuantificador tal como “todo”), dependiendo del contexto del hablante. Pero esto reflejará apenas un hecho parroquial sobre nuestro lenguaje, más que cualquier hecho profundo sobre el mundo. En los aspectos profundos, los objetos virtuales no son menos reales que los objetos ordinarios.

Nota 11: la respuesta a la objeción 8 es alusiva del punto familiar, asociado con Russell y Kant, de que no conocemos la naturaleza intrínseca de las entidades en el mundo externo. Cuando viene a las entidades físicas, la percepción y la ciencia pueden decirnos cómo estas entidades nos afectan, y cómo se relacionan las unas con las otras, pero estos métodos nos dicen poco sobre lo que las entidades físicas fundamentales son en sí mismas. Esto es, estos métodos revelan la estructura causal del mundo externo, pero dejan abierta su naturaleza intrínseca.

La Hipótesis Metafísica es en parte como una hipótesis sobre lo que subyace a esta estructura microfísica causal: las entidades microfísicas están hechas de bits. Lo mismo se aplica para la Hipótesis de la Matrix. Uno podría decir que si estamos en una matrix, ¡la ding-an-sich kantiana (cosa en sí) es parte de una an-sich computarizada! Esta hipótesis complementa nuestra concepción ordinaria del mundo externo, pero realmente no la contradice en la medida en que esta concepción ordinaria permanece en silencio en lo que atañe a la naturaleza intrínseca del mundo.

Nota 12: una moraleja general es que la “imagen manifiesta” es *robusta*: nuestra concepción ordinaria del mundo macroscópico no se ve fácilmente falsada por descubrimientos en ciencia y metafísica. Con tal de que el mundo físico contenga procesos con la clase correcta de estructura causal y contrafáctica, entonces será compatible con la imagen manifiesta. Incluso una simulación computarizada tiene la estructura causal y contrafáctica relevante, al igual que los procesos en la mente de Dios: por esto podemos apoyar una realidad externa robusta, a pesar de su sorprendente naturaleza.

Esta clase de flexibilidad en nuestra concepción del mundo está estrechamente ligada a la no-neutralidad semántica de muchos de nuestros conceptos. Esos conceptos, tales como “agua”, “cabello”, y “electrón”, dejan un poco de flexibilidad con respecto a aquello en que puede convertirse su referente. Concebimos sus referentes aproximadamente como cualquier entidad real que desempeñe un cierto rol causal, o tenga una cierta apariencia, mientras dejamos abierta su naturaleza intrínseca. Uno podría igualmente argumentar que las restricciones más fuertes, impuestas por nuestra concepción del mundo, son plausiblemente aquellas asociadas a los conceptos semánticamente neutrales, los cuales no dan lugar a esta clase de flexibilidad. Estos conceptos incluyen plausiblemente muchos de nuestros conceptos

causales (y nómicos), así como muchos de nuestros conceptos mentales. En estos casos, tenemos una clase de comprensión “directa” de cómo debe ser el mundo para que satisfaga los conceptos. Si esto es así, entonces nuestras creencias causales y mentales imponen restricciones fuertes en la forma como debe ser el mundo real.

Uno puede argumentar que nuestros conceptos fundamentales semánticamente neutrales son conceptos mentales (“experiencia”, “creencia”), conceptos causales (“causa”, “ley”), conceptos lógicos y matemáticos (“y”, “dos”), y conceptos categoriales (“objeto”, “propiedad”). Hay también muchos conceptos semánticamente neutros que involucran más de uno de estos elementos: “amigo”, “acción” y “computadora” son algunos ejemplos. Si esto es correcto, entonces las restricciones fundamentales que nuestras creencias imponen al mundo externo son que éste contenga estados mentales relevantes (en nosotros y otros), y que contenga objetos y propiedades que tienen las relaciones causales relevantes entre sí y con los estados mentales. Esta clase de concepción es lo suficientemente débil para que pueda ser satisfecha por una matrix (por lo menos si ésta es una matrix multi-cubeta, o si el computacionalismo sobre la mente es verdadero).

En mi opinión, este problema sobre las restricciones fundamentales que nuestras creencias imponen al mundo, es el problema filosófico más profundo que surge del pensamiento sobre la matrix. Si lo que he dicho en este artículo es correcto, es precisamente porque estas restricciones son relativamente débiles, que muchas hipótesis que uno pueda haber pensado como “escépticas” resulten ser compatibles con nuestras creencias. Y esto es lo que nos permite construir algún tipo de respuesta al desafío escéptico. Un tanto paradójicamente, uno podría decir que sabemos tanto porque exigimos muy poco.

Nota 13: ¿por qué una simulación computarizada de un mundo satisface estas restricciones? La razón depende de la naturaleza del cómputo y su implementación. Puede considerarse que cualquier cómputo formal ofrece una especificación de una *estructura causal* (abstracta), especificando la manera precisa de interacción entre algunas series de estados formales. Para implementar un cómputo formal de este tipo, se requiere que la implementación tenga estados concretos que realizan directamente esos estados formales, donde el patrón de interacción (causal y contrafáctico) entre estos estados refleje precisamente el patrón de interacción entre los estos formales (ver Chalmers 1994). Así, cualquiera de estas dos implementaciones del cómputo compartirá

una cierta estructura causal específica. Se requerirá una descripción computacional del mundo físico para reflejar su estructura causal hasta el nivel de los objetos y las propiedades fundamentales. Así, cualquier implementación de este cómputo encarnará esta estructura causal (en transiciones entre estados implementadores, sean éstos voltajes, circuitos, o algo completamente diferente). En la medida en que nuestra concepción del mundo externo imponga restricciones a la estructura causal que un mundo físico real pueda satisfacer, estas restricciones también serán satisfechas por una simulación computarizada.

(Esto se relaciona con una observación hecha por Hubert Dreyfus.² Como yo, Dreyfus acepta el punto de vista de que la mayoría de las creencias de los habitantes de una matrix serán verdaderas, no falsas. Pero Dreyfus sugiere que muchas de sus creencias causales serán falsas: e.g. su creencia general en que hay “un universo físico con poderes causales que hacen cosas que suceden en nuestro mundo”, y quizás sus creencias específicas de que los gérmenes causan enfermedades, que el sol causa que las cosas se calienten, y así sucesivamente. Desde mi punto de vista, esta sugerencia es incorrecta. En mi opinión, el mundo de alguien que vive en una matrix tiene causación real que sigue por todas partes dentro de él, basado en la causación real que se da en la computadora. Los gérmenes virtuales en la computadora realmente causan enfermedades virtuales en el computador. Cuando los habitantes en una matrix dicen “los gérmenes causan enfermedades”, lo que dicen es verdadero.)

Por supuesto las restricciones mentales también necesitan ser satisfechas. En particular, es importante que la estructura causal esté en la clase correcta de relación con nuestras experiencias. Pero esta restricción también será satisfecha cuando estemos conectados a una matrix. Las restricciones con respecto a otras mentes serán satisfechas con tal de que estemos en una matrix multi-cubeta, o si el computacionalismo sobre la mente es verdadero. De esta manera, una matrix tiene todo lo que se requiere para satisfacer las cruciales restricciones causales y mentales en nuestra concepción del mundo.

Nota 14: una posible línea de objeción al argumento de este artículo es argumentar algo así como que hay *restricciones adicionales* que nuestras creencias imponen en el mundo, que la Hipótesis de la Matrix no

² El artículo de Hubert Dreyfus (& Stephen Dreyfus) titulado *The Grave New World Of The Matrix*, se encuentra, al igual que el de Chalmers, en la página oficial de Matrix: http://whatisthematrix.warnerbros.com/rl_cmp/phi.html. (N del T)

satisface. Uno puede argumentar que un mero ajuste en la estructura mental y causal no es suficiente. Por ejemplo, uno podría argumentar que el mundo necesita tener las propiedades *espaciales* correctas, donde tengamos alguna clase de contacto directo con lo que son las propiedades espaciales (quizás porque los conceptos espaciales son semánticamente neutrales). Uno puede sugerir que el problema con la matrix es que sus propiedades espaciales son todas erróneas. Creemos que las entidades externas están ordenadas en un cierto patrón espacial, pero tal patrón no existe dentro del computador.

Respuesta, uno puede argumentar que estas restricciones adicionales no existen. Se puede argumentar que los conceptos espaciales no son semánticamente neutrales, sino que en realidad están sujetos a los experimentos mentales de la Tierra Gemela. Mi estudiante Brad Thompson ha desarrollado experimentos mentales de esta clase (Thompson 2003), involucrando una Doble Tierra donde “un metro” se refiere a (lo que llamamos) dos metros, en un Mundo El Greco donde “cuadrado” se refiere a (lo que llamamos) rectángulos, y así sucesivamente. Desde este punto de vista, nuestros conceptos espaciales se relacionan con cualquier clase de propiedades y relaciones en el mundo externo que sean causalmente responsables por nuestro correspondiente conjunto de experiencias espaciales: en este respecto, los conceptos espaciales son análogos a los conceptos de color. Aquí no tenemos ningún contacto “directo” con la naturaleza básica de las propiedades espaciales. En cambio, una vez más, las restricciones básicas son mentales y causales.

Esta línea de objeción es tácitamente revisada en la sección 9 del artículo, donde sugiero que si hay un nivel computacional físico subyacente, entonces cualquier implementación del cómputo formal relevante puede servir, en principio, como una realización de ese nivel, sin comprometer la realidad física. Quizás, un oponente podría negar que pudiera haber un nivel computacional subyacente a lo físico, o por lo menos podría afirmar que hay restricciones acerca de qué tipo de implementación puede servir. Por ejemplo, se podría sostener que el nivel implementado en sí mismo debe tener una estructura espacial apropiada.

Pienso que esta línea de respuesta va en contra del espíritu de la física contemporánea. Los físicos han considerado seriamente la idea de que el espacio como lo comprendemos no es fundamental, pero que hay

un nivel subyacente, no descrito en términos de nociones espaciales ordinarias, del que el espacio emerge. La hipótesis del autómatas celular es justamente una propuesta de este tipo. Aquí, lo que es crucial es simplemente un patrón de interacción causal. Si los físicos descubren que este patrón es realizado, a su vez, por una clase completamente diferente del nivel con propiedades muy diferentes, no podrían concluir que el espacio físico ordinario no existe. Más bien, concluirán que el espacio, en sí mismo, está constituido por algo no espacial. Esta clase de descubrimiento podría ser sorprendente y revolucionaria, pero, de nuevo, no más que la mecánica cuántica. Y como con la mecánica cuántica, casi ciertamente no consideraríamos esto como una hipótesis escéptica sobre el mundo macroscópico externo. Si esto es correcto, entonces nuestra concepción del mundo macroscópico no impone restricciones esencialmente espaciales en el nivel fundamental de la realidad.

Problemas similares surgen con respecto al tiempo. En algún aspecto, el tiempo da menos problemas que el espacio, en la medida en que una simulación computarizada en una matrix se desarrolla en el tiempo, en el mismo orden temporal como el tiempo en el mundo simulado. Así, uno no puede objetar que las disposiciones temporales relevantes no están presentes en la matrix, en la forma que uno puede objetar que las disposiciones espaciales relevantes no están presentes. Aún si los conceptos temporales fueran semánticamente neutrales, la Hipótesis de la Matrix podría aún vindicar nuestras creencias temporales. Todavía pienso que uno podría elaborar un argumento para que nuestro concepto del tiempo externo no sea semánticamente neutral (es notable que los físicos hayan considerado hipótesis en las que las nociones temporales no desempeñan ningún papel en el nivel fundamental). Más bien, esto señala que el conjunto externo de propiedades y relaciones es responsable de nuestro correspondiente conjunto de experiencias temporales. Si esto es así, entonces cualquier simulación computarizada con la estructura causal correcta, y la correcta relación con nuestras experiencias, vindicará nuestras creencias temporales, sin tener en cuenta su naturaleza temporal intrínseca.

Nota 15: el razonamiento en este artículo no ofrece una refutación contundente del escepticismo, en la medida en que muchas hipótesis escépticas quedan abiertas. Pero pienso que fortalece significativamente una de las respuestas habituales dadas al escepticismo. Se sostiene a menudo que, aunque varias hipótesis escépticas son compatibles

con nuestras experiencias, la hipótesis de que hay un mundo físico real provee una más simple o mejor explicación de las regularidades en nuestras experiencias que estas hipótesis escépticas. En este caso, podemos entonces justificarnos al creer en el mundo físico real, por una inferencia hacia la mejor explicación.

A estas alturas se objeta a menudo que algunas hipótesis escépticas parecen tan simples como la explicación habitual: por ejemplo, la hipótesis de que todas nuestras experiencias son causadas por una simulación computarizada, o por Dios. Si fuera así, la anterior respuesta al escepticismo fracasaría. Pero si estoy en lo cierto, entonces estas hipótesis “igualmente simples” no son hipótesis escépticas en absoluto. En este caso, la inferencia hacia la mejor explicación se puede emplear después de todo: todas estas hipótesis “simples” proporcionan principalmente creencias verdaderas sobre un mundo externo.

El problema que persiste concierne a las variadas hipótesis escépticas que permanecen sobre la mesa, tales como la Hipótesis de la Matrix Reciente, la Hipótesis de la Matrix Local, y todas las otras. Sin embargo, parece razonable afirmar que éstas son significativamente menos simples que las hipótesis anteriores. Todas ellas involucran una explicación no uniforme de las regularidades en nuestras experiencias. En la Hipótesis de la Matrix Reciente, las regularidades presentes y pasadas tienen explicaciones muy distintas. En la Hipótesis de la Matrix Local, las creencias sobre asuntos familiares y cercanos, por un lado, y sobre asuntos inusuales y lejanos, por el otro, tienen explicaciones muy distintas. Estas hipótesis, en conjunto, tienen una clase de estructura de mecanismo dual que parece considerablemente más compleja que las anteriores estructuras de mecanismos uniformes. Si esto es correcto, uno puede argumentar que la inferencia hacia la mejor explicación justifica que excluyamos esas hipótesis, y que aceptemos las hipótesis no escépticas.

Incluso si uno piensa que algunas de estas hipótesis escépticas ofrecen explicaciones razonables de nuestra experiencia, todavía queda en el vecindario un prometedor argumento contra el escepticismo global acerca del mundo externo. Si estoy en lo cierto, todas estas hipótesis escépticas son, en el peor de los casos, hipótesis escépticas *parciales*: si son correctas, entonces una buena parte de nuestras creencias empíricas será verdadera, y todavía habrá un mundo externo. Para obtener una hipótesis escéptica *global*, necesitamos ir de todas formas hasta la Hipótesis del Caos. Pero esta es una hipótesis en la que las regularidades

de nuestra experiencia no tienen explicación en absoluto. Incluso una versión extremadamente débil de la inferencia hacia la mejor explicación, justifica que se descarte esta clase de hipótesis. Si esto es así, entonces esta clase de razonamiento puede justificar nuestra creencia en la existencia del mundo externo.

REFERENCIAS

- BOSTROM, N. (2003). Are you living in a computer simulation? *Philosophical Quarterly* 53:243-55. <http://www.simulation-argument.com>.
- CHALMERS, D.J. (1990). *How Cartesian dualism might have been true*. <http://consc.net/notes/dualism.html>.
- _____. (1994). *A computational foundation for the study of cognition*. <http://consc.net/papers/computation.html>.
- DENNETT, D.C. (1978). Brainstorms. In *Where am I?* MIT Press.
- PUTNAM, H. (1975). The meaning of "meaning". In *Mind, Language, and Reality*. Cambridge University Press.
- _____. (1981). *Reason, Truth, and History*. Cambridge University Press.
- SEARLE, J.R. (1984). Can computers think? In *Minds, Brains, and Science*. Harvard University Press.
- THOMPSON, B. (2003). *The Nature of Phenomenal Content*. Ph.D. dissertation, University of Arizona.
- WOLFRAM, S. (2002). *A New Kind of Science*. Wolfram Media.