

PROPIEDADES TELEOLÓGICAS Y SUPERVENIENCIA^{1*}

MANUEL PEREZ OTERO

Departamento de Lógica, Historia y Filosofía de la Ciencia
Universidad de Barcelona

El objeto de este artículo es discutir la tesis de la superveniencia de lo mental sobre lo físico y formular una versión satisfactoria de la misma. Esta tesis afirma que las propiedades mentales "descansan" o *sobrevienen* sobre propiedades (micro)físicas, esto es, que *no puede* haber diferencia en propiedades mentales sin diferencia en propiedades microfísicas. La naturaleza exacta de esta tesis depende de cómo se precise el elemento modal presente en la noción intuitiva de superveniencia; a diferentes modos de precisar el componente modal corresponden diferentes conceptos de superveniencia. El concepto más usual en la literatura, que denominaré F-superveniencia, es demasiado fuerte pues parece haber buenos motivos para dudar que las propiedades mentales F-sobrevengan sobre las físicas. Otros conceptos más débiles, como, por ejemplo, G-superveniencia, parecen, sin embargo, demasiado débiles, pues no discriminan propiedades *prima facie* causalmente eficaces de las que no lo son: las propiedades mentales sí G-sobrevienen sobre las físicas, pero G-sobrevienen además otras que no consideramos causalmente eficaces. En este trabajo presento un nuevo concepto de superveniencia, T-superveniencia, más débil

1. Este trabajo ha sido posible gracias a una beca concedida por el Comissionat per a Universitats i Recerca de la Generalitat de Catalunya. Algunas de las ideas expuestas aquí son fruto de discusiones en un seminario sobre filosofía de la mente coordinado por el profesor Manuel García-Carpintero que se celebró en el Departamento de Lógica, Historia y Filosofía de la Ciencia de la Universidad de Barcelona durante el curso 1992-93. Agradezco a José A. Díez Calzada, Manuel García-Carpintero y David Pineda sus comentarios sobre una versión anterior de este texto que me han ayudado a formular más claramente algunos puntos.

que el primero pero más fuerte que los segundos y que parece un buen candidato a expresar la superveniencia *característica* de lo mental (en realidad, como se defenderá, de cualesquiera propiedades *teleológicas*); por un lado, las objeciones a la F-superveniencia de lo mental sobre lo físico son claramente inválidas respecto de la T-superveniencia; por otro, no aparecen ahora como T-supervenientes sobre lo físico otras propiedades que, a diferencia de las propiedades mentales, coincidiríamos en que carecen de eficacia causal.

En la sección I presento un experimento mental, inspirado en el experimento de la Tierra Gemela de Putnam, que cuestiona la superveniencia de lo mental sobre lo microfísico. En relación con él, se discute una posible réplica según la cual lo que importa para las cuestiones de causación (la eficacia causal de las propiedades mentales está implicada en la discusión) es la superveniencia nomológica o metafísica; esta superveniencia no quedaría refutada apelando a posibilidades meramente conceptuales como lo es (según esta línea de crítica) la Tierra Gemela. A continuación expongo otro experimento mental, tomado de Davidson, el ejemplo del Hombre de los Pantanos, que tiene las mismas consecuencias que el anterior pero que, sin embargo, parece resistir la réplica aludida.

En la sección II preciso cuál es la noción de superveniencia utilizada implícitamente en la discusión precedente; se trata de F-superveniencia o superveniencia fuerte, caracterizada por Kim, que sería violada por las propiedades mentales según muestran los dos experimentos mentales. Tras ello, se presenta otra noción más débil: G-superveniencia o superveniencia global, también de Kim; ésta resulta demasiado débil ya que no da cuenta suficientemente de la mayor eficacia causal que las propiedades mentales parecen tener, en comparación con otras propiedades típicamente ineficaces. Ello pone de manifiesto la necesidad de hallar una noción de superveniencia más fuerte que la superveniencia global pero menos comprometida que la superveniencia fuerte, que sea apropiada a propiedades como las mentales. A lo largo de esta sección

explicito, además, algunos de mis supuestos metafísicos y abordo cuestiones metodológicas sobre la relevancia de la Tierra Gemela y el Hombre de los Pantanos.

En la sección III, finalmente, utilizo la elucidación del concepto de función de Wright para definir la noción de propiedad teleológica; asumo que las propiedades intencionales son un tipo de propiedades teleológicas e introduzco la nueva noción de superveniencia (T-superveniencia) que tiene las virtudes exigidas en la sección anterior.

I

1. El primero de los experimentos mentales a que me he referido es una variación del de Putnam [75] que estaba destinado a mostrar que los significados de algunas palabras no quedan completamente determinados por lo que sucede en el *interior* de las cabezas de quienes las usan. Se nos pide que consideremos dos personas, Oscar y BiOscar, habitando, respectivamente, la Tierra y una Tierra Gemela, tan similar como queramos a nuestro planeta, con la salvedad de que en vez de agua hay un líquido de apariencia indistinguible pero estructura atómica diferente, XYZ. BiOscar es un duplicado de Oscar: habla un lenguaje como el inglés, excepto que su término "agua" refiere a XYZ, ha tenido una historia similar y sus creencias sobre el *agua* son *exactamente las mismas* que las de Oscar sobre el agua.² Dado que de acuerdo con nuestras intuiciones el término "agua" no significa lo mismo en ambos lenguajes (por ser la identidad de la extensión un requisito razonable de ello) concluimos que los *estados internos* (lo que ocurre *dentro* de sus cabezas) de los individuos no bastan para determinar el significado de un término; en casos como éste

2. Así es como lo expresa Putnam, aunque, como señalaré enseguida, no considero correcto afirmar que sus creencias sean las mismas puesto que las creencias de BiOscar no son, en rigor, sobre el agua (otros criterios de individuación de estados psicológicos sí identificarían las creencias de Oscar y BiOscar; no voy a desarrollar aquí, sin embargo, las razones por las que tales criterios me parecen inadecuados). Por ello, habría que precisar de otro modo la noción de *estado interno* necesaria para establecer el argumento, y respecto a la cual Oscar y BiOscar no difieren.

factores externos también cuentan. Para nuestros fines, que son discutir la superveniencia no de los significados sino de los estados mentales sobre estados físicos, añadamos a lo estipulado por Putnam el que BiOscar sea un duplicado microfísico idéntico de Oscar, en el sentido, vago, de que ambos satisficieran exactamente las mismas descripciones de una supuesta teoría física fundamental correcta, o, al menos, de una teoría física correcta no menos básica que ninguna de las que se están investigando hoy (de modo correspondiente habrá de entenderse mi uso, en adelante, de "propiedades microfísicas" y expresiones similares). Tenemos así que, aunque esto contradiga el planteamiento literal que hacía Putnam del experimento, el mismo tipo de intuiciones que nos impedían identificar los dos significados de "agua" en el ejemplo inicial nos muestran que tampoco los estados mentales de Oscar y BiOscar son el mismo: las interrelaciones entre significado y creencia hacen que sea incorrecto atribuir a BiOscar estados mentales como creer que hay agua a su izquierda. Y el resultado del experimento es doble: por una parte, las propiedades mentales no sobrevendrían sobre propiedades microfísicas de los individuos que las ejemplifican, es decir, podría haber dos individuos que compartiendo todas sus micropropiedades difirieran en cuanto a alguna de sus propiedades mentales (en la sección siguiente seré más preciso sobre la noción de superveniencia); por otra parte, y como consecuencia de lo anterior, si el que una propiedad sobrevenga sobre alguna propiedad microfísica es un requisito necesario para ser causalmente eficaz (es decir, para que los objetos que la posean tomen parte en relaciones causales singulares en virtud de poseer la propiedad en cuestión) entonces las propiedades mentales no serían causalmente eficaces.³ Quien considere implausibles tales consecuencias tiene aquí un problema.

3.Estoy usando "propiedad" en un sentido muy amplio, sin pretender incluir únicamente propiedades *naturales*.

2. Una de las soluciones sugeridas, que asume el carácter extrínseco de las propiedades mentales sin renunciar a la tesis de la superveniencia de lo mental sobre lo microfísico, aparece en García-Carpintero [1993]. Podemos discernir ahí dos líneas de reivindicación diferentes de la eficacia causal de las propiedades mentales. La principal de ellas será abordada en las dos secciones siguientes; la otra, que se encuentra también en García-Carpintero [92], consiste en señalar que la noción de superveniencia que interesa a la cuestión de la causación (en este caso a la causación de lo mental) es la superveniencia nomológica o metafísica, para refutar la cual sólo es lícito apelar a circunstancias o mundos nomológica o metafísicamente posibles.⁴ Son esos mundos los que determinan la extensión de propiedades como contrapuestas a los conceptos: un caso paradigmático de la distinción: las expresiones "oro" y "Au" refieren a una misma propiedad (ver Kripke [72] y Putnam [75], cuyas ideas sobre el modo de referir de los términos de *géneros naturales* asumo como básicamente correctas) pero expresan diferentes conceptos; siendo las condiciones de individuación de los conceptos, por lo general, más estrictas que las de las propiedades, la clase de los mundos conceptualmente posibles incluirían, en principio, a la clase de los mundos nómicamente posibles (es decir, habría, por ejemplo, mundos conceptualmente posibles con objetos que son de oro pero no tienen la estructura Au, aunque ninguno de tales mundos sería nómicamente posible).⁵ Se afirma, entonces, que la situación de la Tierra Gemela es

4. En la sección II, §5 doy una breve justificación de por qué acepto esa identificación de lo nomológicamente posible con lo metafísicamente posible.

5. Estoy asumiendo un modo *realista* de hablar sobre propiedades (e incluso sobre conceptos) pero creo que la distinción que hacemos aquí entre propiedad y concepto asociados a un predicado puede trazarse desde concepciones diferentes (cuando sea factible en absoluto, pues no descarto que en muchos casos esa discriminación sea imposible). Conviene señalar, además, que representar la relación entre la diferente fuerza modal de propiedades y conceptos mediante un conjunto de mundos metafísicamente posibles incluido propiamente en un conjunto de mundos conceptualmente posibles no es más que una aproximación bastante limitada.

una circunstancia sólo conceptualmente pero no nómicamente posible: en vez de la relación "oro"/"Au" tendríamos ahora, por ejemplo, "es una creencia de que hay agua a su izquierda"/una expresión de la teoría física fundamental que refiera a la presunta micropropiedad, por compleja que sea, correspondiente a tal tipo de creencia.⁶

Parecería que esa salida no es completamente satisfactoria; pues aunque ciertamente el experimento mental de la Tierra Gemela no es, en principio, nada más que eso, una posibilidad mental, conceptual, tiene todo el aspecto de ser algo más: está construido de un modo relativamente limpio y legal con ingredientes poco sospechosos de anomicidad; si, pese a ello, no hay Tierras Gemelas en ningún mundo nomológicamente posible, ¿qué tipo de leyes (¿de qué nivel: geológicas, biológicas, ...?) son las que prohíben la existencia de una Tierra Gemela? ¿es quizá que biológicamente no puede haber individuos muy similares a nosotros pero sin H₂O en su organismo? Pero si sólo fuera eso podría restaurarse el problema sustituyendo "agua" por "aluminio".

3. Un modo de responder a esa dificultad haciendo más plausible la posición de García-Carpintero sería atender, primero, a casos de estados mentales que involucren conceptos de un contenido más básico que el de agua, conceptos observacionales como *rojo*. Teniendo en cuenta que un requisito para atribuir a alguien la posesión de ese concepto será que tenga una disposición a detectar

6.En aras de la claridad he hecho la analogía tal vez más cerrada de lo que sea correcto. Es razonable pensar que no haya una expresión en la teoría microfísica básica correspondiente a propiedades como creer que hay agua a la izquierda. Debido a la múltiple realizabilidad de las propiedades mentales, es probable que se correspondan con disyunciones infinitas de micropropiedades (o de propiedades definidas a partir de micropropiedades y de propiedades postuladas en teorías cuasi-empíricas sobre formación de objetos compuestos a partir de objetos simples). Esto no obsta al núcleo de la discusión aunque genere sus propias controversias.

una determinada propiedad de los objetos de su entorno, la de ser rojo, y que tal disposición debe tener alguna implementación neurofisiológica y, en última instancia, microfísica, parece entonces bien razonable creer que las leyes de la física (o de la física y de la biología) impongan suficientes restricciones al tipo de estructura microfísica nómicamente correlacionada del modo adecuado con la rojez (restringido, a su vez, por otras condiciones, por ejemplo la de formar parte de estructuras más complejas que sean organismos vivos) como para que ese mismo tipo de estructura no pueda tener la misma correlación con otra propiedad distinta como, por ejemplo, la amarillez, es decir, no pueda ser también un concepto de *amarillo* (más razonable cuanto más dispares resulten microfísicamente rojez y amarillez). Para extender posteriormente esa ilustración a conceptos más complejos que rojo, por ejemplo a agua, aluminio o artritis, haría falta, primero, disponer de un análisis de los mismos que mostrase su dependencia con respecto a otros conceptos más observacionales. Así, la imposibilidad nomológica de que haya individuos microfísicamente idénticos pero mentalmente diferentes se heredaría de abajo a arriba, entre los diferentes estratos de nuestro sistema conceptual (ver García-Carpintero [93], pg. 15). Aunque todo eso sea muy esquemático bastaría para dudar de que la Tierra Gemela fuera una circunstancia nómicamente posible.

4. Sea cual sea su relevancia con respecto a la Tierra Gemela, las reflexiones que acabo de apuntar no sirven, sin embargo, para afrontar nuestro segundo experimento mental. También en esta ocasión se postula una réplica microfísica de un individuo normal y corriente, Donald (de Donald Davidson en el ejemplo original, ideado por él en Davidson [86]) pero formado por pura coincidencia como resultado de la caída de un rayo sobre un árbol muerto en una zona pantanosa. Esta vez el duplicado, el Hombre de los Pantanos, carece por completo de estados mentales, carece, en

particular, de los que quepa atribuir a Donald. No es difícil ver la razón de que el Hombre de los Pantanos sea inmune a las consideraciones de I, §3 contra la posibilidad nomológica de la Tierra Gemela: tales consideraciones utilizaban lo exigentes que son los requisitos a satisfacer por quienquiera que posea estados intencionales para sugerir que BiOscar, dotado de intencionalidad por estipulación del experimento, podría estar más allá de lo físicamente permitido por las leyes; pero ninguna restricción similar pesa sobre la naturaleza del Hombre de los Pantanos, pues éste se origina azarosamente; ese origen puede tomarse como un medio *ad hoc* de liberar al Hombre de los Pantanos de cualesquiera propiedades (aparte de las estrictamente implicadas por su disposición microfísica) que, al constituirle, limitaran su naturaleza tal y como ocurría con BiOscar.

Creo que debe admitirse que el Hombre de los Pantanos es algo más que una mera posibilidad conceptual permitida por la distancia entre nuestros conceptos mentales y microfísicos; representa una posibilidad microfísica genuina, si bien altamente *improbable* dado lo altamente compleja que debe ser la estructura microfísica tipo correspondiente a un ser intencional. Para reforzar lo que estoy sosteniendo introduzco esta variación sobre el ejemplo: sea no la de Donald sino la de un martillo la estructura física duplicada al caer un rayo en una zona pantanosa. Ahora la macropropiedad que distingue al Martillo de los Pantanos del martillo original es la de ser un martillo; y la estructura de un martillo puede ser tan simple como para resultar desencaminado negar posibilidad física al Martillo de los Pantanos.⁷ Resulta, entonces, que una propiedad

7. Naturalmente, uno puede estar tentado a reaccionar ante este ejemplo (e incluso ante el del Hombre de los Pantanos) negando que el duplicado carezca de la macropropiedad en cuestión. Ante eso sólo cabe invocar nuestros juicios intuitivos, con la esperanza de que éstos no se vean enturbiados por versiones tal vez demasiado fuertes de principios filosóficos por sí mismos muy plausibles como el fiscalismo o por carecer, de momento, de una conjetura filosófica que permita articular con cierta precisión un tratamiento más ecuánime de esos juicios. Espero

como ser martillo no sobreviene sobre las propiedades microfísicas de sus ejemplificaciones; no debería extrañarnos que ocurriera lo mismo con propiedades filosóficamente menos inocuas, como las propiedades mentales.

Todo esto sugiere que la asimilación de los casos de la Tierra Gemela o el Hombre de los Pantanos a la posibilidad meramente conceptual no es correcta. Pese a ello, me parece que el tratamiento adecuado de esos casos consiste en explicar por qué hay un sentido en que cabe discriminarlos como irrelevantes; un sentido relacionado con la otra línea de defensa de la eficacia causal de la mente ejercida en García-Carpintero [93]: mostrar la crucial diferencia que hay entre las propiedades mentales y algunos ejemplos paradigmáticos de propiedades causalmente ineficaces. Antes de abordarlo (en la sección III) convendría explicitar y discutir el concepto de superveniencia; me ocupó de ello en la siguiente sección.

II

5. La noción de superveniencia que he venido usando hasta ahora fue caracterizada por Jaegwon Kim (en Kim [1984a] y Kim [1984b]) en los siguientes términos:

La familia de propiedades A sobreviene fuertemente (F-sobreviene) sobre la familia de propiedades B si y sólo si, necesariamente, para cada x y cada propiedad J en A, si x tiene J, entonces hay una propiedad K en B tal que x tiene K, y necesariamente si algún y tiene K, entonces tiene J.⁸

que la elucidación del concepto de función (en III, §8) constituya tal conjetura.

8. Kim la denomina superveniencia fuerte porque la contrasta con la superveniencia débil, definida exactamente igual pero sin el segundo operador de necesidad, con lo que se permite que en cada mundo posible la propiedad G mencionada sea una diferente. La superveniencia de las propiedades semánticas de un lenguaje sobre sus propiedades sintácticas es el único ejemplo que conozco de superveniencia débil sin superveniencia fuerte. Conservaré el calificativo, sin embargo, pues también sirve para contraponer esa noción de superveniencia a otras que voy a examinar y

Las clases A y B están cerradas bajo las operaciones booleanas de complementación, conjunción, disyunción y conjunción infinita de propiedades.⁹ A será la clase F-superveniente (y las propiedades en A propiedades F-supervenientes) y B la clase de F-supervenencia. Cuando tomamos como A la clase de todas las macropropiedades y como B la de todas las micropropiedades, denominamos superveniencia mereológica a la tesis general de que A F-sobreviene sobre B (Kim [1984b], pg. 261) (cuando alguna familia de propiedades bajo discusión, D, contenga una única propiedad, E, hablaré de superveniencia de, o sobre, E queriendo significar superveniencia de, o sobre, D; espero que el contexto impida cualquier confusión).

Tal y como señala Kim, se obtienen diferentes tesis de superveniencia para clases de propiedades fijadas, A y B, dependiendo de cuál sea la modalidad que atribuyamos al operador "necesariamente" que ocurre doblemente en la definición. Asumiré, como se apuntaba en I, §2, que no es la necesidad conceptual la que nos interesa sino la necesidad metafísica o nomológica; y ésta última me la represento, en términos muy esquemáticos, del siguiente modo: creo que los mundos metafísicamente posibles son aquéllos que es necesario postular para atestiguar cualquier combinación *posible* de posesión o carencia de propiedades (incluyendo todo tipo de propiedades) por parte de cualesquiera objetos, estando determinadas tales combinaciones por las compatibilidades e incompatibilidades entre propiedades, derivadas de la naturaleza de cada propiedad (entiendo "propiedad" en un sentido suficientemente realista como para no individualizar propiedades epistemológica-

que, ciertamente, son más débiles.

9. Aunque a Kim le pasa inadvertido, por el contexto de la discusión en Kim [1984a] se sobreentiende que deben admitirse conjunciones infinitas de propiedades, salvo que se modifique la definición.

mente, aunque sin implicar "propiedad causalmente eficaz"). Con ese esbozo en perspectiva, podríamos entonces expresar ciertas tesis sobre relaciones entre las posibilidades combinatorias de las micropropiedades y las posibilidades combinatorias de las macropropiedades como tesis sobre existencia o inexistencia de mundos posibles. Así, por ejemplo, Kim denomina microdeterminismo a la tesis de que no hay dos mundos posibles que sean microfísicamente idénticos (Kim [1984b], pg. 261): en caso de haberlos sería porque no serían idénticos respecto a alguna propiedad no microfísica; el microdeterminismo afirma precisamente que eso es imposible, que *dos* mundos microfísicamente idénticos no difieren tampoco respecto a ninguna otra propiedad, siendo por tanto un único mundo.¹⁰ Acepto el microdeterminismo como verdadero, si bien no voy a intentar justificarlo aquí. Utilizando la premisa razonable de que la variabilidad microfísica no sobrepasa lo establecido por las leyes microfísicas (aunque razonable, discutida), el microdeterminismo permite concluir que los mundos metafísicamente posibles son exactamente los permitidos por las leyes microfísicas. Por esa razón asimilo lo metafísicamente posible con lo nomológicamente posible (*nomológico* con respecto a las leyes microfísicas, que serían, en caso de haber otras, las más permisivas).

En nuestros experimentos mentales la clase presuntamente F-superveniente era la clase de las propiedades que adscribimos habitualmente mediante las atribuciones de estados mentales de la psicología popular, "es una creencia de que p", "es un deseo de que

10.El microdeterminismo debe ser distinguido de las nociones usuales de determinismo causal que el nombre puede evocar. El determinismo causal puede caracterizarse como la tesis de que no hay dos mundos posibles idénticos en todos los respectos hasta algún momento de tiempo, *t*, pero que difieren en algo con posterioridad a *t*.

p", "es una percepción de x".¹¹ No es tan inmediato identificar la presunta clase de F-superveniencia; observemos que para que la Tierra Gemela y el Hombre de los Pantanos se acrediten como contraejemplos a la F-superveniencia de lo mental sobre lo microfísico, este *lo microfísico* ha de entenderse comprendiendo sólo las propiedades microfísicas *intrínsecas* de los individuos, aquéllas cuya ejemplificación no depende conceptualmente de lo que ocurre externamente al cuerpo del individuo (ver Horgan [1982], pg. 33). Se supone que sólo ese tipo de propiedades vendrían relacionadas por las leyes últimas de la microfísica, y, por lo tanto, que así debería entenderse "micropropiedades" en nuestra definición, más arriba, de superveniencia mereológica (en el caso del Hombre de los Pantanos esa localización espacial iría acompañada de una localización temporal puesto que el Hombre de los Pantanos sólo es microfísicamente idéntico a Donald con posterioridad a la caída del rayo, antes no existía). Esto suscita dos reflexiones. La primera es que hay muchas macropropiedades que no parecerían amenazar una postura filosófica fisicalista simpatizante con la superveniencia mereológica, pero que, sin embargo, claramente no F-supervienen sobre propiedades microfísicas en ese sentido: estar a menos de 1 Km. de distancia de un caballo o estar quemado por el sol (mi piel quemada por el sol puede ser microfísicamente indistinguible de la piel de otra persona que se quemó de otra manera y que, por lo tanto, no está quemado por el sol; por ello, estar quemado por el sol no F-sobreviene sobre ninguna micropropiedad (este segundo ejemplo está en Davidson [1986], pg. 143)) son muestras de ello. La segunda reflexión, al hilo de la anterior, es que cabe preguntarse si no sería deseable disponer de una elucidación diferente de la relación de superveniencia que

11. Asumo, pues, que el tipo de entidades que ejemplifican las propiedades mentales no son personas sino estados, tratando éstos como una categoría de objetos.

permitiera asertar una tesis general de superveniencia más débil que la superveniencia mereológica que, a diferencia de ésta, no fuera flagrantemente violada por macropropiedades tan simples como las dos recientemente citadas; al fin y al cabo, lo que no resulta problemático es señalar qué es lo que impide a BiOscar y al Hombre de los Pantanos tener los estados mentales de Oscar y Donald, respectivamente: el que no mantengan ciertas relaciones causales históricas sistemáticas con elementos de su entorno; tal vez habría que ampliar la clase de superveniencia complementando las micropropiedades intrínsecas con un determinado tipo de (no todas ellas) propiedades espaciales e históricas de modo que se recogieran selectivamente esos factores extrínsecos. Por desgracia, formular con precisión una idea así no resulta sencillo y, además, en la medida en que me represento vagamente cómo podría funcionar, creo que casos análogos al Hombre de los Pantanos volverían a mostrar que se violaba la superveniencia que se obtuviera. Hay que contentarse con algo más modesto y optar por una definición diferente.

6. Siguiendo, todavía, a Kim (Kim [1984a], pg. 168), digamos que dos mundos son discernibles con respecto a la familia de propiedades A si para alguna J en A y un individuo x, x tiene J en uno pero no en el otro (en caso contrario son indiscernibles); definamos ahora:

A sobreviene globalmente (G-sobreviene) sobre B si y sólo si mundos indiscernibles con respecto a B (B-indiscernibles) son también A-indiscernibles.¹²

12. Kim afirma que G-superveniencia es equivalente a F-superveniencia pero lo cierto es que la primera es más débil que la segunda; hay un error en su derivación de la F-superveniencia a partir de la G-superveniencia: que exista un mundo posible en que un objeto tiene la propiedad B' (de entre la clase de propiedades B) pero no tiene la propiedad F no implica que para algún objeto x de un mundo posible w' en que F(x) y B'(x) haya un mundo, w'', B-indiscernible de w', en que $\neg F(x)$ (ver Kim [1984a], pg. 168).

A y B son familias de propiedades con las mismas condiciones que antes.

También ahora podemos generalizar: siendo A la clase de todas las propiedades y B la clase de las micropropiedades intrínsecas (en el sentido antes indicado) obtenemos la tesis que avanzamos más arriba con el título de microdeterminismo (tomando "microfísicamente idénticos" como "microfísicamente indiscernibles").¹³

Es claro que ni el Hombre de los Pantanos ni la Tierra Gemela son contraejemplos a la G-superveniencia de lo mental sobre lo microfísico. En el caso de la Tierra Gemela, por ejemplo, los mundos posibles en que están Oscar y BiOscar son, ciertamente, psicológicamente discernibles pero también son microfísicamente discernibles.

Y no podía ser de otro modo ya que, como asumí anteriormente, la tesis general microdeterminista es verdadera: las propiedades mentales, la de estar quemado por el sol, cualquier propiedad, en definitiva, por más extrínseca que sea, está microfísicamente determinada de la manera en que lo afirma el microdeterminismo.

13.El microdeterminismo es una tesis muy similar a la que resulta si interpretamos la superveniencia mereológica bajo una lectura *no local*: aceptando entre las micropropiedades también las extrínsecas en el sentido más amplio posible, de modo que la historia microfísica entera de un mundo se adscribiera *leibnizianamente* a cada uno de sus objetos mediante una o más propiedades. Terence Horgan (Horgan [1982], pg. 33) indica que es posible una interpretación así de la superveniencia mereológica, pero lamenta que es demasiado débil para capturar la noción intuitiva de determinación de lo macrofísico por lo microfísico (así como bajo la interpretación *local* esa tesis era violada por aquellas propiedades que, como las intencionales, son dependientes del contexto, de factores externos al individuo que las ejemplifica (Horgan [1982], pg. 32)); por ello, Horgan propone una formulación afín que pretende recoger la idea de la determinación de cualquier macropropiedad por lo que microfísicamente ocurre dentro de un individuo o dentro de una porción relativamente local del universo (Horgan [1982], pgs. 33-34), utilizando para ello las nociones de rasgo *intrínseco* a y rasgo *cualitativo* de una región espacio-temporal. Logra formular, entonces, una tesis de superveniencia que aparentemente (sólo aparentemente, en mi opinión) es algo más fuerte que el microdeterminismo.

Pero eso no es satisfactorio. Lo que defienden muchos filósofos es que las propiedades mentales son causalmente eficaces, que un determinado estado físico puede afectar y ser afectado causalmente en virtud de ser un estado con contenido mental.¹⁴ Incluso si se admite la no F-superveniencia de lo mental sobre lo microfísico desearíamos un criterio clasificador más fino que el que engloba las propiedades mentales con cualquier propiedad por ineficaz que sea. Necesitaríamos, quizá, una noción de superveniencia más fuerte que el microdeterminismo pero más débil que superveniencia mereológica que se ajustara a las peculiaridades de propiedades como las mentales.

Y la razón de ello es muy afín a la que presenta García-Carpintero (García-Carpintero [1993], pgs. 11-13): hay una radical diferencia entre las propiedades mentales y otras de cuya eficacia causal típicamente desconfiamos, como viajar en Marte 13 o haber nacido bajo la influencia de Marte (o estar a menos de 1 Km. de distancia de un caballo, estar quemado por el sol, etc.): existe un vínculo explicativo entre las propiedades mentales y las propiedades microfísicas que se estén ejemplificando y sean genuinamente responsables en cada caso de presunta causación mental; un vínculo que podemos hallar y exhibir pero que es inverosímil que exista en el caso de las otras propiedades supuestamente eficaces. Tomemos el caso de la propiedad haber nacido bajo la influencia

14. Entre esos filósofos no se encuentra Davidson, que no creo que sostuviera nunca nada más fuerte que el microdeterminismo (en Kim [1984a] se insinúa lo contrario). Tres datos apoyan esa afirmación: (1) en Davidson [1986], pg. 125, tras idear el Hombre de los Pantanos, asume sin ningún problema que éste carece de intencionalidad; (2) en Davidson [1974], pgs. 252-253, para justificar, entre otros ejemplos, por qué un gesto que sea de asentimiento en Austria y de disentimiento en Grecia no deja de sobrevenir físicamente afirma que basta incrementar el marco físico de referencia para encontrar una diferencia relevante: puesto que Austria es físicamente distinta de Grecia, cualquier evento en Austria es físicamente distinto de un evento en Grecia; (3) el microdeterminismo es bastante coherente con lo anómalo del monismo defendido por Davidson.

de Marte, llamémosla M_m , y un presunto efecto de la misma: ser agresivo, A_m . Si creemos que con respecto a A_m la eficacia causal de M_m es nula es porque nos resulta inverosímil que la micropropiedad (altamente extrínseca) correspondiente a M_m , llamémosla M_r (que consistiría, *grosso modo*, en estar en ciertas relaciones espaciales con determinados objetos, planetas, en un determinado momento de tiempo, el del nacimiento) tenga la más mínima relación con la micropropiedad, CA_r , causalmente responsable de la aparición de la micropropiedad ejemplificada cada vez que se ejemplifica A_m ; con lo cual, no hay, presumiblemente, ninguna relación entre M_m y CA_r . Pero el panorama es diferente cuando consideramos las macropropiedades mentales: aunque voy a mantener que la micropropiedad correspondiente a una macropropiedad mental no es la misma que la micropropiedad causalmente responsable del macroefecto pertinente, lo cierto es que sí sospechamos que hay una relación entre esta última micropropiedad y la macropropiedad mental en cuestión, un vínculo explicativo que sugiere que la G-superveniencia no agota, no describe exhaustivamente los modos en que lo mental *descansa* sobre lo microfísico.

En la siguiente sección intentaré contribuir a la identificación de ese vínculo y caracterizaré el concepto de superveniencia que estábamos buscando.

7. Ahora que disponemos de un mayor bagaje terminológico quisiera, antes de retomar, en la siguiente sección, el hilo principal de la argumentación, hacer un breve paréntesis para intentar situar en su justa medida el *status* y la relevancia de nuestros dos experimentos mentales. Lo que se está discutiendo es si la clase, A, de propiedades mentales sobreviene sobre la clase, B, de propiedades microfísicas (intrínsecas). Sea K la conjunción de las propiedades microfísicas de un determinado estado concreto de Donald, e, el correspondiente a una percepción de rojo, por ejemplo; y sea J la propiedad mental de ser una percepción de rojo. Es cierto que sólo la existencia en el mundo real de un estado con

la propiedad K pero que carezca de la propiedad J puede servir para establecer concluyentemente que J no sobreviene fuertemente sobre K; pues quien defiende que sí se da tal superveniencia niega que el Hombre de los Pantanos sea nomológicamente posible y puede, con toda coherencia, mantener su postura salvo que se le muestre que es real (el modo más crudo de mostrar que algo es posible). Sin embargo, el experimento mental del Hombre de los Pantanos (y análogamente en el caso de la Tierra Gemela) no constituye una petición de principio, no es una estipulación de que lo mental no F-sobreviene sobre lo microfísico puesto que su estructura no es "consideremos un estado concreto, e', microfísicamente idéntico a e (que posee, pues, K) pero que no es una percepción de rojo (que carece, pues, de J)". Su estructura, más bien, es la siguiente: "consideremos un estado concreto, e', microfísicamente idéntico a e (que posee, pues, K) y que posee también la propiedad microfísica (extrínseca) K' (donde K' sería, *grosso modo*, ser un estado microfísico de un objeto que en el momento t, anterior al actual, fue configurado aleatoriamente a partir de las moléculas de un árbol al sufrir, éste, una descarga eléctrica)". Sólo eso se estipula en el experimento mental. Que un estado que sea K' no puede ser J es algo en lo que (muchos) concordamos posteriormente, que se deriva de nuestro concepto de J (en general, de nuestros conceptos de lo mental). Lo que el experimento mental presupone es que K y K' son propiedades microfísicamente compatibles. Quien mantiene la F-superveniencia de J sobre K quiere defender que alguna ley psicofísica correlaciona K con J. Pero entonces, dado que conceptualmente K' y J son incompatibles, debe defender que está microfísicamente prohibido ser K y K'. La fuerza del experimento del Hombre de los Pantanos reside en que nos parece falso (a quienes lo aceptamos) que alguna combinación de leyes microfísicas pueda resultar en una incompatibilidad de K y K' (fuerza que es menor en el caso análogo para la Tierra Gemela, exactamente por las razones aducidas en I, §3).

III

8. El ejemplo del martillo no era casual. La propiedad de ser un martillo, como otras muchas propiedades (ser un corazón, ser un televisor, etc.), es una propiedad ejemplificada por un objeto si y sólo si ese objeto tiene una determinada función. Denominemos propiedades teleológicas a las que son de ese tipo.¹⁵ El concepto intuitivo de función, tanto de función natural como de función artificial, fue definido por Larry Wright (Wright [1973]) mediante las siguientes cláusulas:

La función de X es Z si y sólo si

- (1) X hace Z,
- (2) X existe porque hace Z, y
- (3) Z es resultado de que X exista.

No me extenderé sobre las bondades de esta definición, que la convierten, a mi parecer, en un paradigma de análisis filosófico; simplemente voy a asumir que es correcta.

El objeto de traer esto a colación es que la línea de investigación y análisis del contenido de los estados mentales que considero más prometedora es la que atribuye a éstos funciones (en ese sentido, teleológico, de Wright), postulando la posesión de una función como un requisito necesario para que un estado neurofisiológico constituya un estado mental (así lo vienen defendiendo Ruth Garret Millikan y Fred Dretske). Aceptando ese enfoque, podemos ver con más detalle por qué, por ejemplo, el Hombre de los Pantanos no posee estados mentales. Consideremos el estado microfísico X correspondiente a un estado mental cualquiera atribuible a Donald: percepción de rojo, pongamos por caso. Lo que ese estado *hace*, Z, será, simplificando, interactuar con otros

15. Evito usar "propiedades funcionales" ya que podría ser equívoco, dado que esa expresión suele utilizarse con otro significado en filosofía de la mente.

estados microfísicos de Donald (correspondientes, algunos de ellos, a otros estados mentales) para finalmente causar en él cierto tipo de movimientos, los apropiados a la presencia de un objeto rojo. Pues bien, pese a que el *token* de X en el Hombre de los Pantanos efectivamente hace Z (se cumple la cláusula (1) de la definición de función) y Z es el resultado de que exista ese *token* de X (se cumple (3)), resulta que Z no es una función de X ya que no se satisface la cláusula (2): ese *token* de X no existe porque haga Z, sino por azar; por eso el estado microfísico X del Hombre de los Pantanos no es también un estado mental. Lo mismo sucede con el Martillo de los Pantanos: sea cual sea la función Z de los martillos, el Martillo de los Pantanos no existe porque haga Z; ésa es la razón de que no sea un martillo.¹⁶ La crucial cláusula (2) de la definición de Wright marca la diferencia entre una función y una mera disposición, es el rasgo característico de las propiedades teleológicas.¹⁷

9. La definición de función permite identificar el vínculo explicativo entre las macropropiedades mentales y las propiedades microfísicas al que aludíamos en II, §6 (y que existe también para cualesquiera propiedades teleológicas): la existencia de objetos (o

16.Sugiero que la creencia de que el Martillo de los Pantanos sí es un martillo, como los demás casos en que una propiedad teleológica, R, es ejemplificada por un objeto en virtud de tener una función artificial (es decir, de ser un artefacto) podría ser indebidamente propiciada por el hecho de que en general no nos interesamos en si un objeto es o no es un R sino en si *funciona como* o *puede servir como* un R, y para esa noción de *funcionar como* un R sí que bastan las cláusulas (1) y (3) de la definición que especifique qué es tener la función propia de un R.

17.Aunque no es una de las varias propuestas que Wright examina antes de dar con su definición, el análisis que ofrece Ernest Nagel de los enunciados teleológicos en Nagel [1961] incluye una condición prácticamente equivalente a esa cláusula (2) de Wright. Es notorio, sin embargo, que Nagel no es consciente de la importancia de esa condición pues la discusión consecuente procede ignorándola completamente, incluso hasta el punto de tratar presuntas objeciones a su formulación que lo son debido sólo a esa desatención (ver Nagel [1961], cap. XII, pgs. 403 y siguientes).

estados) con la macropropiedad es explicada por la presencia de la micropropiedad. En efecto, el estado mental X existe porque hace Z (lo leemos en la cláusula (2)) y, dado que ese hacer Z es explicado, en último termino, por la propiedad microfísica que X esté ejemplificando, tenemos que X existe porque ejemplifica la propiedad microfísica en cuestión.

Ese vínculo, por cierto, es a priori: supuesto que el que un concepto corresponda a una propiedad teleológica es analítico, es decir, que el análisis teleológico de un concepto (entendiendo por "análisis teleológico de un concepto" el análisis que indica que el concepto se aplica a un objeto si y sólo si éste posee una determinada función) es un caso de análisis conceptual, es conceptualmente necesario, y supuesto que la definición de función de Wright (mediante la cual se han caracterizado las propiedades teleológicas) también lo es (en García-Carpintero [1992], nota 20, hay una concisa refutación de lo contrario) entonces quien comprende el concepto de martillo (o de ser una creencia de que p) sabe, a priori, que existe ese vínculo explicativo entre la macropropiedad martillo y las micropropiedades que en cada caso se estén ejemplificando; si saber esto último basta para saber que la macropropiedad es causalmente eficaz (como parece sugerirse en García-Carpintero [1993], pg. 15) entonces la eficacia causal de ciertas propiedades, al menos las teleológicas, está implicada en los conceptos correspondientes a ellas. Lo extraña que resulta esa conclusión apoya la aserción que haré en IV, §11 de que cabe interesarse por otros sentidos más fuertes de eficacia causal, como el correspondiente a F-superveniencia.

Estoy ya en disposición de introducir una nueva noción de superveniencia (diferente de las que formuló Kim: F-superveniencia y G-superveniencia) que resultará apropiada a las propiedades teleológicas:

La familia de propiedades A T-sobreviene sobre la familia de propiedades B si y sólo si, necesariamente, para cada x y cada

propiedad J en A, si x tiene J, entonces hay una propiedad K en B tal que x tiene K, y necesariamente si x tiene K, entonces tiene J. Que no es más que la definición de F-superveniencia de Kim substituyendo "algún y" por "x".

Afirmo que es apropiada a las propiedades teleológicas porque éstas, si bien violan superveniencia mereológica (como atestiguan el Hombre de los Pantanos y el Martillo de los Pantanos), sí T-sobrevienen sobre propiedades microfísicas intrínsecas (cosa que no ocurre con haber nacido bajo la influencia de Marte o estar quemado por el sol). He aquí una demostración informal de ello: sea x un objeto en un mundo posible, w, que tiene la propiedad teleológica J, siendo su función Z; K será la conjunción infinita de propiedades microfísicas que x posee en

algún momento en que ejemplifica J. Supongamos, por reducción al absurdo, que hay un mundo posible, w', en que x tiene K pero no J. Asumo como cierta la premisa de que la parte disposicional de J (encarnada en las cláusulas (1) y (3) de la definición de Wright) sí T-sobreviene (incluso F-sobreviene) sobre G. Así pues, es por incumplir la cláusula (2) por lo que x no tiene J en w': en w' x existe pero no existe porque haga Z. Pero eso contradice que x exista en w porque hace Z puesto que si un objeto existe porque p entonces en cualquier mundo metafísicamente posible en que exista existe porque p (este principio, aunque razonable, constituye la parte más delicada de la prueba).

Tomemos otra vez el estado microfísico X que se corresponde con el estado mental de percibir rojo de Donald. Sí existen casos concretos de ese estado microfísico tipo X que no se corresponden con el estado mental (en el Hombre de los Pantanos); pero el caso concreto de estado X en Donald, en toda circunstancia posible en que exista existirá porque haga Z (a través de una compleja historia evolutiva); luego en toda circunstancia posible ese *token* de X satisface los tres requisitos para que Z sea su función, es decir, es

un estado con el contenido mental correspondiente a tener la función Z.

10. Conviene ahora discernir dos sentidos diferentes en que, aun admitiendo que los mundos de la Tierra Gemela y el Hombre de los Pantanos fueran microfísicamente posibles, se les podría considerar irrelevantes.

El primero estriba en afirmar que es la T-superveniencia lo que importa; ciertamente, esos mundos no son contraejemplos a la T-superveniencia de lo mental sobre lo físico, dada la definición de T-superveniencia.

El otro consiste en sostener que tales posibilidades con respecto a las leyes microfísicas serían imposibilidades con respecto a ciertas leyes no básicas que involucran macropropiedades y que sólo lo nómicamente posible con respecto a estas últimas leyes es lo que importa dado que macropropiedades son las que están en cuestión (ver García-Carpintero [1992], nota 25); imagino que el modo en que esto segundo podría ocurrir concierne al carácter falible de esas macroleyes que prohibirían casos como el Hombre de los Pantanos debido a lo muy improbable de su existencia, mientras que las microleyes sí lo admitirían (de modo análogo a cómo una macroley puede prohibir que una mano a 37° C de temperatura que se sumerja durante 10 segundos en agua hirviendo mantenga al final la misma temperatura aunque eso sea microfísicamente posible). Tal vez, efectivamente, muchas macropropiedades no F-sobrevengan sobre lo microfísico debido sólo a esas posibilidades microfísicas *raras que son imposibilidades* con respecto a macroleyes; ello sugiere que quizá lo adecuado sea hallar un modo no *ad hoc* de especificar cómo la amplia clase de mundos microfísicamente posibles debe restringirse, en cada caso en que se considera una macropropiedad, a mundos posibles *apropiados* al macronivel que corresponda a esa macropropiedad; en caso de lograrse, la presunta superveniencia sobre lo microfísico

de esa macropropiedad podría plantearse, no mediante la noción de T-superveniencia, sino mediante una variación de la noción de F-superveniencia consistente en que el operador de necesidad que aparece en ella sólo tome en cuenta los mundos posibles determinados por esa especificación. No dudo que debiera emprenderse esa tarea pero rechazo una idea que tal estrategia podría suscitar: la idea de que es intrínseco a la distinción entre *macro* y *micro* el que ninguna macropropiedad F-sobrevenga (en el sentido original, sin restringir los mundos posibles) sobre propiedades microfísicas. Propiedades como agua, oro o rojo sí F-sobrevienen sobre lo microfísico por muy admisivos que seamos respecto a los mundos microfísicamente posibles. En efecto, consideremos un experimento análogo al del Hombre de los Pantanos: "supongamos que como resultado de la caída de un rayo sobre un árbol muerto se forma un objeto de estructura subatómica Au, pero que no es de oro". Respuesta: imposible. Pero no porque sea nomológicamente imposible que un rayo reestructure un árbol y forme Au, se admite eso por improbable que sea. Lo que ocurre es que dadas nuestras intuiciones sobre el modo de referir de términos como "oro", en esas circunstancias ya queda fijado que el objeto resultante tiene la propiedad de ser de oro. La réplica análoga en el caso del Hombre de los Pantanos por parte de quien considera a éste una posibilidad meramente conceptual habría de ser: "admito que un rayo reestructure un árbol en forma de un objeto microfísicamente idéntico a Donald. Lo que ocurre es que dadas nuestras intuiciones sobre el modo de referir de términos como 'percepción de rojo', en esas circunstancias ya queda fijado que el estado resultante tiene la propiedad de ser una percepción de rojo". Pero nuestras intuiciones son justamente las contrarias: en esas circunstancias (y porque "percepción de rojo" o "martillo" no refieren del mismo modo que "oro", "calor" o, incluso, "rojo") ya queda fijado que el estado resultante no es una percepción de rojo. Así pues, dado que sí existen macropropiedades cuya dependencia sobre lo microfísico

es tan fuerte como lo exigido por F-superveniencia, no es ocioso indagar en cada caso si así ocurre con una determinada macropropiedad que estemos inspeccionando; en particular, los dos experimentos mentales mostrarían que las propiedades intencionales no exhiben una dependencia tan fuerte sobre lo microfísico.

11. Las conclusiones que deseo extraer de la discusión son las siguientes. Hay al menos tres nociones (o grados) diferentes y relevantes de superveniencia metafísica entre familias de propiedades (y diferentes de la superveniencia conceptual): G-superveniencia, T-superveniencia y F-superveniencia. Ante cada propiedad o clase de propiedades filosóficamente interesante tiene perfecto sentido preguntarse si sobreviene, y en qué forma de superveniencia, sobre alguna otra propiedad o propiedades. Si es plausible llamar fisicalismo (o materialismo) a un principio filosófico que defienda que toda propiedad sobreviene sobre propiedades microfísicas intrínsecas, hay tres versiones, de diferente fuerza, del fisicalismo, correspondientes a las tres lecturas posibles de sobrevenir. Igualmente, si vinculamos conceptualmente la eficacia causal de una propiedad a su superveniencia sobre lo microfísico, hay, en general, tres grados o tipos de eficacia causal que cabe atribuir a las propiedades (tal vez siendo el grado más débil, el correspondiente a mera G-superveniencia, un caso extremo, de nula eficacia causal). Experimentos mentales como el del Hombre de los Pantanos (y, quizá, el de la Tierra Gemela) han servido para ver que las propiedades mentales no F-sobrevienen sobre propiedades microfísicas intrínsecas. Como propiedades teleológicas que es razonable creer que son, sólo la parte disposicional de las propiedades mentales tiene toda la eficacia causal que se puede tener porque sólo esa parte F-sobreviene sobre lo microfísico; las propiedades mentales, como tales, sólo T-sobrevienen, y ése es el tipo de eficacia causal que tienen. Ese no es un rasgo característico de la distinción entre micropropiedad y macropropiedad, ya que

hay macropropiedades, como oro, caliente o rojo, que (por el modo de significar de los predicados con los que usualmente nos referimos a ellas) sí F-sobrevienen sobre las micropropiedades, mostrándose, así, que la distinción entre F-superveniencia y T-superveniencia está lejos de ser irrelevante.

REFERENCIAS

- DAVIDSON, Donald [1974]: "The Material Mind", D. Davidson, *Essays on Actions and Events*, Oxford: Clarendon Press, 1982, pgs. 245-259.
- [1986]: "El conocimiento de la propia mente", Carlos Moya comp. y trad., *Donald Davidson. Mente, mundo y acción*, Barcelona: Ed. Paidós, 1992, pgs. 119-152; original publicado en *Proceedings and Adresses of the American Philosophical Association*.
- GARCIA-CARPINTERO, Manuel [1992]: "Dretske on the Causal Efficacy of Meaning", presentado en el IV seminario interdisciplinar de ciencia cognitiva en Sagunto, enero 1992; próxima aparición en *Mind and Language*.
- [1993]: "The Supervenience of Mental Content", presentado en el V seminario interdisciplinar de ciencia cognitiva en Barcelona, mayo 1993; próxima aparición en *Proceedings of the Aristotelian Society*.
- HORGAN, Terence [1982]: "Supervenience and Microphysics", *Pacific Philosophical Quaterly*, 63, 1982, pgs. 29-43.
- KIM, Jaegwon [1984a]: "Concepts of Supervenience", *Philosophy and Phenomenological Research*, 45, 1984, pgs. 153-176.
- [1984b]: "Epiphenomenal and Supervenient Causation", David M. Rosenthal ed., *The Nature of Mind*, Oxford: U.P., 1991, pgs. 257-265.
- KRIPKE, Saul [1970]: *Naming and Necessity*, Harvard: Harvard U.P., 1980.
- NAGEL, Ernest [1961]: *The Structure of Science: Problems in the Logic of Scientific Explanation*, New York: Harcourt, Brace & World, 1961.
- PUTNAM, Hilary [1975]: "The meaning of 'meaning'", H. Putnam, *Mind, Language and Reality*, Cambridge: U.P., 1975, cap. 12.
- WRIGHT, Larry [1973]: "Functions", *Philosophical Review*, 82, 1973, pgs. 139-168.