

SOBREDETERMINACION CAUSAL MENTE-CUERPO

(*Mind-Body Causal Overdetermination*)

Agustín VICENTE*

Manuscrito recibido: 1998.12.11.

Versión final: 1999.5.25.

* Departamento de Lógica y Filosofía de la Ciencia, Universidad del País Vasco/Euskal Herriko Unibertsitatea, Apartado 1.249, 20080 Donostia-San Sebastián. E-mail: ylbvibea@sf.ehu.es

BIBLID [0495-4548 (1999) 14: 36; p. 511-524]

RESUMEN: Jaegwon Kim ha actualizado y resumido el problema cartesiano de la causalidad mental en tres ideas en conflicto: el principio del cierre causal del mundo físico, la eficacia causal de la mente, y el principio de exclusión causal-explicativa (PEE). Este último principio nos dice que no puede haber dos causas/explicaciones causales que sean ambas *completas e independientes* para un evento determinado, salvo en casos de sobredeterminación. Aunque la forma habitual de afrontar este problema de exclusión es buscar una relación de dependencia entre las propiedades físicas y las mentales, algunos filósofos mantienen que puede tratarse de un caso de sobredeterminación. En este artículo, analizo la posibilidad de que esto sea así.

Descriptores: exclusión causal, superveniencia, fisicismo, sobredeterminación, análisis contrafáctico.

ABSTRACT: *Jaegwon Kim has very nicely updated and summed up Descartes' problem of mental causation in three conflicting ideas: the principle of the causal closure of the physical, the causal efficacy of the mental, and the principle of the causal-explanatory exclusion (PEE). This last principle tells us that there cannot be two causes/causal explanations that are both complete and independent for one event, except in cases of overdetermination. Though the usual way to this exclusion problem is look for a dependency relation between mental and physical properties, some philosophers hold it can be a case of overdetermination. In this paper, I analyze the chances that this could be so.*

Keywords: *causal exclusion, supervenience, physicalism, overdetermination, counterfactual analysis.*

Jaegwon Kim ha resumido y actualizado el problema cartesiano de la causalidad mental de una forma eficaz y concisa en las siguientes tres proposiciones aparentemente inconsistentes:

(i) principio del cierre causal del mundo físico: todo evento físico causado tiene un antecedente causal igualmente físico (para eventos como ins-

tanciaciones de propiedades: toda instanciación de una propiedad física causada es causada por la instanciación de otra propiedad física);

(ii) eficacia causal de la mente: los eventos mentales producen cambios en el mundo físico;

(iii) principio de la exclusión causal-explicativa (PEE): no puede haber dos causas/explicaciones causales que sean tanto completas como independientes para un evento, salvo en los casos de sobredeterminación,

y la siguiente cuestión: ¿cómo, dados (i) y (iii), puede ser (ii) verdadera? A este problema se le ha dado en llamar el *problema de la exclusión causal-explicativa*.

Voy a suponer la verdad de (i) y (iii), pero cabe motivar y discutir ambas proposiciones brevemente. En primer lugar, y especialmente si somos fisicistas, tendremos mucha simpatía por (i), ya que, según se dice, se trata de una ley que se desprende de nuestra física actual: en palabras de Antony y Levine (1997, p. 33), "[n]ada puede afectar la distribución de materia y energía en el espacio-tiempo salvo la instanciación de propiedades básicas en los objetos básicos que ocupan el espacio-tiempo". Es además de suponer que este principio posee una cierta fuerza modal; en particular, que rige en todos aquellos mundos en los que hay propiedades físicas en absoluto.

En segundo lugar, (iii) parece simplemente resumir (acertadamente) algunas de nuestras prácticas explicativas: como norma general, cuando se nos ofrecen dos explicaciones con apariencia de independientes, bien deshacemos tal apariencia, bien nos quedamos con una de ellas, mientras rechazamos la otra. Alternativamente, cabe entender que (iii) enuncia una verdad conceptual: cabe decir que define la sobredeterminación.

Si entendemos que nuestro tercer enunciado no es una definición, sino un resumen de nuestras prácticas explicativas, cabe entonces afirmar que hay filósofos que intentan solucionar el problema de la exclusión explicativa negando el principio de exclusión PEE (por no ser un buen resumen de tales prácticas, por ejemplo). No obstante, si leemos el PEE como una definición de la sobredeterminación, entonces tales filósofos no estarían negando la tercera premisa del problema de la exclusión -no pueden negar una definición- sino proponiendo resolverlo mediante el recurso a la sobredeterminación. Al fin de acotar y no complicar en exceso la discusión, pues las posturas de estos filósofos están llenas de diversos matices y tienen un alcance mayor del que pretendo ahora cubrir, propongo entender el PEE de la primera manera, de modo tal que los autores a los que hago referencia

vendrían a mantener que entre las explicaciones mentales y las explicaciones físicas no existe competencia causal, i.e. el PEE no se aplica. Los defensores de la sobredeterminación que contemplaremos más adelante, en contraste, aceptarían la vigencia del PEE, observarían por lo tanto un conflicto causal entre propiedades mentales y propiedades físicas, pero lo resolverían diciendo que se trata de un caso de sobredeterminación causal.

Así pues, tiene sentido decir que hay quienes han rechazado el (PEE), y que también hay quienes simplemente dicen haberlo hecho. Así, Tyler Burge (1993), Terence Horgan (1997) y Scott Sturgeon (1998) han negado su validez, al menos en lo que concierne a propiedades que compiten causalmente desde distintos niveles. El caso de Horgan es especialmente llamativo: reconoce que el PEE está fuertemente atricherado en nuestras intuiciones, y sin embargo, rechaza que pueda existir exclusión causal cuando las propiedades en litigio pertenecen a distintos niveles explicativos. Afirma entonces que aplicar el PEE al caso mente-cuerpo es un error, y trata de explicar por qué se da ese error que tan arraigado tenemos en nuestra intuición. Su idea es que la percepción de una exclusión entre propiedades mentales y propiedades físicas es una "ilusión cognitiva" del tipo de la ilusión Muller-Lyell, una ilusión que, aunque sepamos que es tal, no podemos evitar que nos someta. Sin embargo, creo que es sencillo ver que hay más asimetrías que parelismos entre ambos tipos de "ilusión": en especial, uno puede medir las dos líneas de la ilusión Muller-Lyell y conocer así la realidad, mientras que cabe preguntarse cómo puede llegarse a un conocimiento equivalente en metafísica. Desde luego, no a través de las intuiciones, pues es una intuición la que sostiene al PEE.

Como digo, hay otros filósofos (William Lycan puede ser un ejemplo) que afirman rechazar el principio de exclusión causal-explicativa, pero que, por así decirlo, se traicionan al proponer una relación entre mente y cuerpo, o, en general, entre las propiedades de un nivel y las de niveles inferiores. Esta es, de hecho, la manera más común de afrontar el problema de la exclusión: buscar una relación de dependencia entre las propiedades mentales y las propiedades físicas tal que el PEE no pueda afectar a la causación mental. Recordemos que el PEE nos dice que no puede haber dos causas ambas completas e independientes. Esto quiere decir que podemos tener causación mental si la hacemos dependiente de la causación física (o viceversa).

Esta forma de afrontar el problema de la exclusión se ha explotado durante algunos años, siendo la superveniencia local la relación de dependencia favorita para hacerlo. Si lo mental superviene localmente sobre lo

físico, se ha dicho, entonces lo mental tiene eficacia causal (véase Kim (1984), Fodor (1987)). Por otra parte, si las propiedades de niveles superiores no son capaces de supervenir localmente sobre lo físico, como es el caso de los contenidos amplios, entonces carecen de potencia causal (véase, de nuevo, Fodor (1987)). Por consiguiente, la superveniencia local era tenida por condición tanto necesaria como suficiente para la eficacia causal de las propiedades de niveles superiores.

No obstante, la noción de superveniencia ha resultado no ser demasiado útil, debido a su oscuridad y falta de poder explicativo -aparte de ser incapaz de resolver lo que se ha llamado "el problema del *qua*" (ver Horgan (1989), (1992))- Hay, por esta razón, un deslizamiento reciente hacia la relación de realización en este frente; esto es, parece que los filósofos no están satisfechos con la noción de superveniencia y buscan en el concepto de realización un sustituto adecuado a fin de rehuir el problema de la exclusión. El que lo puedan obtener o, por el contrario, descubran que la relación de realización es esencialmente reductiva (véase Kim (1998)), es algo que no nos va a ocupar ahora.

Me propongo explorar otro modo de resolver el problema de la exclusión causal-explicativa. El PEE nos dice que no pueden existir dos causas para un evento salvo en los casos de sobredeterminación. En lo que sigue, quiero analizar la posibilidad de que causación mental y causación física estén determinando doble e independientemente la instanciación de algunas propiedades físicas, una línea de ataque al problema de la exclusión que apenas se ha explorado, aunque muchas veces haya sido objeto de amago.

1. En primer lugar: ¿cómo debe entenderse la afirmación de que los comportamientos están sobredeterminados causalmente? Tenemos algunos casos prototípicos de sobredeterminación causal, como el de el pelotón de fusilamiento que efectúa dos disparos mortales simultáneos contra un prisionero, o el de la cerilla que prende un pajar al mismo tiempo que el rayo que cae sobre él. En esos casos -tomemos el del pelotón de fusilamiento- tenemos un par de contrafácticos que sabemos verdaderos:

(1) si uno de los disparos no hubiera matado al prisionero, lo habría hecho el otro;

(2) si este otro disparo no hubiera matado al sujeto, entonces lo habría hecho el primero.

Podemos llamar a esta forma de entender las afirmaciones de sobredeterminación causal el "análisis clásico". Si aplicamos este análisis al caso de la causación mental, tendremos:

(1') si (la instanciación de) la propiedad mental M no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad física F lo habría causado;

(2') si (la instanciación de) la propiedad física F no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad mental M lo habría causado.

Se supone que tanto (1') como (2') rigen para cualquier M, F y B tales que M y P se coinstancian en el mundo real y anteceden (presumiblemente causalmente) a B. Por otra parte, se entiende que B es una propiedad física. Tal vez no sea adecuado asimilar comportamientos y eventos físicos, pero tal cuestión no es relevante en este momento. Quien tenga reparos, en cualquier caso, puede pensar que "B" está por movimientos corporales.

Algunos (véase, e.g. Block (1990)) han puesto objeciones a esta forma de entender la sobredeterminación mente-física, basándose en que de ella se desprenden coincidencias inexplicadas y constantes. Los casos de sobredeterminación causal son más bien escasos, por lo que la afirmación de que casi todos los eventos están sobredeterminados causalmente suena, al menos en principio, realmente extraña. El dibujo del mundo que este sencillo análisis traza -supuesto que ampliamos nuestra "solución" al problema de la causación mental a todas las ciencias especiales- es el de una cantidad ingente de propiedades de distintos niveles, algunas de las cuales comparten poderes causales y resultan coinstanciarse (casi) cada vez que se da un evento en nuestro mundo.

De este modo, si la causación mente-física es un caso de sobredeterminación, ha de ser de una suerte diferente. En particular, la coinstanciación de M y F no debe ser una cuestión de pura coincidencia.

2. Block (1990) sugiere que este puede ser el caso. Quizás podemos aceptar la sobredeterminación sin comprometernos con un mundo hecho de coincidencias; si podemos hacer eso, entonces tal vez estemos en la senda de la solución al problema de la exclusión. Tim Crane (1995) va un poco más lejos que Block, y ofrece una explicación de por qué no es una coincidencia el que las propiedades físicas y las mentales se coinstancien cuandoquiera que es necesario: hay leyes psicofísicas.

Supongamos, por ejemplo, que las propiedades mentales supervienen sobre las propiedades físicas: los enunciados de superveniencia nos dicen que si dos sujetos instancian las mismas propiedades físicas, entonces instancian las mismas propiedades mentales. De esta forma, tendríamos una explicación para la coinstanciación de propiedades mentales y propiedades físicas. Y lo mismo ocurre con las relaciones de dependencia mereológica, realización, determinable-determinado, etc. Si se diera cualquiera de estas relaciones entre las propiedades mentales y las propiedades físicas, entonces tendríamos la explicación de cómo éstas y aquéllas sobredeterminan los comportamientos.

Nótese, sin embargo, que si esto es así, entonces (1') y (2') no son verdaderos. En su lugar tenemos

(1'') si (la instanciación de) la propiedad mental M no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad física F tampoco lo habría hecho;

(2'') si (la instanciación de) la propiedad física F no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad mental M podría o no haberlo causado.

(1'') y (2'') encuentran justificación en el enunciado de superveniencia. Suponiendo que los mundos más cercanos en los que M no causa B son aquellos en los que M no se da, tenemos que, como M superviene sobre F, F tampoco se da, i.e. F tampoco causa B, i.e. (1'') es verdadero. Por la misma razón, cuando F está ausente, M puede darse o no darse, y así (2'') también es verdadero supuesta la superveniencia de M sobre F.

De modo que parecemos disponer de una mejor noción de la sobredeterminación mente-física: esa es la buena noticia. La mala es que ni la queremos ni la necesitamos. Si hay superveniencia (y la superveniencia es, al cabo, antirreduccionista) entonces tenemos asegurada la causación mental desde el comienzo: no hay necesidad de recurrir a la cláusula de la sobredeterminación. Volvamos al PEE: el principio nos dice que dos causas son compatibles sólo en tanto (a) una complete a la otra, o (b) una dependa de la otra; finalmente, el PEE admite una excepción: la sobredeterminación.

De esta manera, si tenemos superveniencia o cualquiera de esas relaciones que pueden basar un enunciado de superveniencia, contamos con una relación entre las dos historias causales y no existe necesidad alguna de desarrollar una nueva noción de sobredeterminación. Sin embargo, habíamos abandonado esa línea de ataque al problema de la exclusión. Es decir,

tenemos que suponer que no hay tal relación. La cuestión acerca de la posibilidad de que los comportamientos estén causalmente sobredeterminados sólo tiene sentido bajo esa suposición.

3. Así pues, parece que tenemos que regresar a nuestra primera formulación de la idea de que la causación mente-física es un caso de sobredeterminación, esto es, al análisis clásico:

(1') si (la instanciación de) la propiedad mental M no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad física F lo habría causado;

(2') si (la instanciación de) la propiedad física F no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad mental M lo habría causado.

Dejando de lado la naturaleza milagrosa de un mundo que hace verdaderos a ambos contrafácticos, propongo entrar en un cierto detalle y ver si existe alguna posibilidad de que resulten ambos verdaderos.

El primer contrafáctico expresa la suficiencia de las causas físicas para sus efectos. Se trata de algo que se sigue de la premisa (i) más arriba, esto es, el principio del cierre causal de la física. No importa el que una propiedad mental se instancie o no: cualquier efecto físico tiene una causa física. En este caso, esa causa física es aquella ligada a B por una ley del mundo real, esto es, F.

Ciertamente, (1') tiene una consecuencia contraintuitiva, a saber, que en mundos cercanos al nuestro hay individuos que se comportan igual que nosotros pero que carecen de propiedades mentales en absoluto. (1') quiere decir "habría ido al armario a por una aspirina aunque no hubiera tenido el dolor" (Crane, 1995, p. 232). Pero nótese que esta consecuencia es contraintuitiva sólo bajo la suposición de que lo mental se relaciona con, o depende, de un modo u otro, de lo físico.

El segundo contrafáctico es bastante más controvertido.

Kim, por ejemplo, cuando toma en consideración posibles soluciones a su problema de la exclusión, rechaza la de la sobredeterminación basándose en la falsedad de (2'): si los eventos conductuales estuvieran causalmente sobredeterminados, entonces (2') tendría que ser verdadero, y tal cosa es contraria al principio del cierre causal del mundo físico.

Robert Van Gulick (1993) y LePore y Loewer (1987), por otra parte, lo creen verdadero:

(...) en el mundo más cercano en el que Sosa desea saciar su sed pero su deseo no es un [F], aún causa que extienda su mano. (L&L, p. 639)

Propongo explorar, a continuación, qué tipo de explicación sustenta esta afirmación y cuáles son sus implicaciones.

Comencemos con lo que dicen LePore y Loewer: afirman, si M se instanciara, pero F no, entonces aún habría una conexión causal de M a B. ¿Cómo se sabe esto? ¿Qué es lo que garantiza la existencia de esa conexión?

Su respuesta es: "si *c* hubiera sido un deseo de saciar la sed pero no hubiera sido [F], entonces habría tenido alguna otra propiedad [F*]" (p. 639). Haciendo las traducciones necesarias de eventos davidsonianos a estructurados: si M se hubiera instanciado sin que lo hubiera hecho F, entonces habría habido una F* que se habría instanciado.

Los supuestos subyacentes a esta afirmación parecen ser: (a) cuandoquiera que se instancia una propiedad mental, hay alguna propiedad física que se instancia; (b) las conexiones psicológicas y psicofísicas se mantienen constantes a través de los mundos físicamente diferentes más próximos al nuestro.

¿Cuál es la evidencia de (a) y cuál la de (b)? (a) puede entenderse como una tesis de realización. De acuerdo al realizacionismo, todas las propiedades de niveles superiores tienen que tener una "teoría de la realización" (Poland (1994)), i.e. han de estar realizadas por un mecanismo físico. Esto volvería a implicar la existencia de una relación entre lo mental y lo físico, y, consecuentemente, la negación de la autonomía para las propiedades mentales, si no fuera por (b).

(B) nos dice que las conexiones psicológicas y psicofísicas y, en general, las conexiones entre propiedades de niveles superiores son preservadas a través de (algunos) mundos físicamente diferentes. Dada nuestra explicación de (a), una forma natural de entender esto es diciendo que las propiedades de niveles superiores "se las apañan" para encontrar realizadores en aquellos mundos en los que el realizador de este mundo no se instancia. Quiere esto decir que puede haber un sentido en el que las propiedades de niveles superiores son *autónomas*, no dependientes.

(Una forma menos interesante de entender (b) es: las propiedades de niveles superiores están realizadas por propiedades físicas complejas como $F_1+F_2+\dots+F_j+\dots+F_n$. Entonces, hay un mundo posible muy cercano en el que Fj no se instancia, pero tanto la propiedad de nivel superior y el efecto (comportamiento o lo que sea) lo hacen, *porque la ausencia de Fj no marca ninguna diferencia*. Si llamamos F a $F_1+F_2+\dots+F_j+\dots+F_n$ y F* a

$F_1 + F_2 + \dots + (\sim F_j) + \dots + F_n$, entonces la afirmación de LePore y Loewer se convierte en verdadera. Asumiremos que el cambio de F a F^* es sustantivo).

De modo que el dibujo, finalmente, queda de la siguiente guisa: las propiedades mentales y de nivel superior necesitan realizadores a fin de desarrollar algún trabajo causal; no obstante, no los "esperan" pasivamente. Las conexiones psicológicas y psicofísicas, y las propiedades implicadas en ellas, *reclutan* realizadores. Tal cosa ocurre porque, en palabras de Van Gulick, "se auto-mantienen o auto-reproducen ante fuerzas físicas perturbadoras que podrían degradarlas o destruirlas" (p. 251). Puede sonar extraño, pero sólo tenemos que pensar en el hecho bastante común de una parte del cerebro que realiza funciones que corresponden a otra parte que ha sido herida.

Este modelo de relaciones entre niveles superiores y niveles inferiores podría llamarse "modelo del equipo de baloncesto": los equipos de baloncesto necesitan realizadores, no pueden existir sin cinco individuos en la cancha. Sin embargo, hay un sentido en el que los equipos de baloncesto no dependen de sus jugadores para conseguir lo que consiguen: en los mundos posibles más cercanos en los que esos jugadores no juegan en ellos, esos equipos aún existen y consiguen muchas de las cosas que consiguen en este mundo. Y tal cosa ocurre no sólo porque son múltiplemente realizables, sino -y básicamente- porque tienen la "intención" de autopreservarse.

Podemos entonces decir que hemos dado con una lectura razonable de (2') más arriba:

(R) si (la instanciación de) la propiedad física F no hubiera causado el comportamiento B , entonces (la instanciación de) la propiedad mental M lo habría causado por medio de una F^* que puede hacerse que produzca B .

Por lo tanto, la cuestión clave es: ¿cómo se las apañan las propiedades de niveles superiores para encontrar realizadores físicos? Primero, ¿los crean? La opción de la creación implica a negación del principio del cierre causal, principio que rige, al menos, en todos aquellos mundos en los que la física relevante es como la nuestra. Además, no queremos negarlo: no en vano, todo esto empezó porque queremos resolver el problema que este principio (junto con el PEE) suponía para la causación mental.

La idea de Van Gulick es que las propiedades de niveles superiores no alteran las leyes de la física; tan sólo reclutan y *activan selectivamente* poderes causales. Algunos entrenadores de fútbol y baloncesto parecen compartir las ideas de Van Gulick; así, bajo la concepción del fútbol de Louis

Van Gaal, los equipos cogen a los jugadores y activan sus poderes de acuerdo a un esquema funcional muy restringido. Pero también los cerebros dañados funcionan de esa manera: F, que realiza a M, deja de existir o cumplir su rol causal; M es una propiedad a preservar bastante importante, lo mismo que su conexión con B, de tal modo que el cerebro se reajusta, y F*, que comparte ciertos poderes causales con F, es reclutado para causar B. Van Gulick: "Es debido a la existencia y persistencia del patrón por lo que los constituyentes concretos de sus instancias fueron reclutados y organizados como están" (p. 252)

¿Es esta idea compatible con el principio del cierre causal del mundo físico? Bueno, la reorganización de un cerebro, un equipo de fútbol o lo que sea, es un cambio físico. De acuerdo a nuestro principio, por lo tanto, ha de tener una causa que es también física: no puede ser mental, a menos que haya sobredeterminación, pero si decimos que tal es el caso, entonces estamos a punto de comenzar una regresión.

Así pues, parece que si alguien quiere comprometerse con (2') a través de (R), tiene que rechazar el cierre causal de lo físico. Por lo tanto, (2') no puede ser verdadero.

4. Antes de entrar en la parte final de este análisis, cabe echar un vistazo a una teoría -la de Dretske- capaz también de sustentar un enunciado semejante a (2').

La teoría dretskeana sobre la causación mental nos dice que las propiedades mentales causan estructurantemente lo que las propiedades físico-neurológicas causan desencadenantemente. Imaginemos -dice Dretske- que un terrorista pone una bomba en el coche de un general, bomba que sólo hará explosión en el momento en que se accione la llave de contacto. Tal cosa ocurre, finalmente, una semana después de colocada la bomba, y es entonces cuando el coche salta por los aires. Si nos preguntamos por la causa de la explosión, cabe dar dos respuestas: (a) el coche explotó porque se accionó la llave de contacto, y (b) el coche explotó porque un terrorista puso una bomba. Ambas respuestas son compatibles -sostiene Dretske- porque apuntan a causas de distinto cariz, desencadenante la primera, y estructurante la segunda, tipos de causas que epistémicamente podemos diferenciar porque el primero nos proporciona información del "cómo" y el segundo del "por qué" del evento a explicar.

La causación mental, entonces, cae dentro del patrón de la causación estructurante: ¿cómo ocurre esto? Supongamos una estructura C en un individuo que tiene ciertas propiedades relacionales, en concreto, que es capaz

de indicar (i.e. ser causada por) la propiedad F del entorno. Supongamos asimismo que poseer un detector de Fs es de considerable importancia para el individuo que tiene C, ya que su supervivencia depende de que ningún F-instanciador le atrape. Entonces, puede ocurrir que, debido a sus propiedades indicadoras, C sea *reclutado* para causar el movimiento corporal B, comportamiento de huida. De esta forma, cada vez que aparece un F-instanciador en las proximidades del poseedor de C, este individuo efectuará el movimiento de huida B. Y la causa de B será (a) desencadenantemente, la instanciación/activación de ciertas propiedades físicas de C; (b) estructurantemente, la instanciación, previa al proceso de reclutamiento, de las propiedades indicadoras de C.

Creo factible afirmar que este esquema teórico podría soportar un enunciado como (2'), (recordemos: (2') "si (la instanciación de) la propiedad física F no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad mental M lo habría causado"), ya que, como en el caso de la teoría realizacionista, cabe suponer que

(D) si (la instanciación de) la propiedad física F de C no hubiera causado el comportamiento B, (dado que era de cierta importancia para el individuo) entonces (la instanciación de) la propiedad indicadora I lo habría causado (estructurantemente) igualmente por medio del reclutamiento de alguna F* que pudiera ser conectada causalmente con B.

De esta manera, cabe decir que la teoría de Dretske puede defender (2') por medio de (D), un enunciado bastante semejante a (R) más arriba.

Los problemas que encuentra esta solución, por otra parte, son equivalentes a los que acuciaban al realizacionismo de Van Gulick y LePore y Loewer: el llamado proceso de reclutamiento es un proceso físico que, de acuerdo al principio de cierre causal del mundo físico, ha de tener un antecedente causal físico (salvo sobredeterminación, etc.); comprometerse con (D), tal como está, supone entonces negar nuestro principio, cosa que hemos acordado no podemos hacer.

5. Parecen quedar, sin embargo, dos opciones, aunque algo desesperadas, por explorar:

(1) negar el principio del cierre causal de lo físico: llamemos a ésta la "opción prohibida". Incluso aunque no estuviera prohibida, las críticas que se aplican a la segunda opción se aplican a esta.

(2) suponer que las propiedades mentales (o indicadoras) son físicas (contra este "abuso o trivialización del término 'físico' véase Poland (1994)). Esta segunda opción, como vamos inmediatamente a ver, es un callejón sin salida, ya que (a) establece una relación entre la causación mental y la causación física, y (b) convierte a (1') en falso.

Muy brevemente: (2') puede ser verdadero y compatible con el cierre causal del mundo físico si y sólo si las propiedades mentales resultan ser propiedades físicas. El problema de la exclusión, bajo ese supuesto, se convierte en un problema "de familia": tenemos dos propiedades físicas F y M compitiendo por un mismo efecto B, y un PEE que nos dice que tenemos que quedarnos con una de ellas, salvo (a) que haya con una relación entre ambas, o (b) haya sobredeterminación.

Tanto (R) como (D), en las dos explicaciones que hemos desarrollado, dicen que hay una relación entre la historia causal mental y la física: F* es reclutada para realizar M y preservar su conexión con B (o, simplemente, para causar B), esto es, F* está ahí, y causa B, *porque* M tiene que estar ahí y causar B (o, porque hay un individuo instanciador de F* que también instancia una propiedad indicadora).

Dejando ahora de lado a Dretske, cuyo caso es aún más claro, creo que no es raro decir que "fue así desde un principio" (aunque, en cualquier caso, podríamos motivar un problema de exclusión para M y F*), esto es, que F fue también reclutada como un realizador de M porque podía conectarla a B. Siendo esto así, parece que la única manera en que (2') puede ser verdadera es negando la tesis de la sobredeterminación (como algo distinto de la "tesis de la relación").

Además, esta teoría hace a (1') falso. Como hemos visto, la conexión F-a-B depende de la conexión M-a-B. Por lo tanto,

-(1') si (la instanciación de) la propiedad mental M no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad física F tampoco lo habría hecho.

6. Para terminar: parece existir una sola forma de defender la sobredeterminación mente-física: quitarle al principio del cierre causal toda fuerza modal. Si pudiéramos hacer tal cosa, entonces sí que podría ocurrir:

(1') si (la instanciación de) la propiedad mental M no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad física F lo habría causado;

(2') si (la instanciación de) la propiedad física F no hubiera causado el comportamiento B, entonces (la instanciación de) la propiedad mental M lo habría causado.

La verdad de (1') nos la daría no el principio de cierre causal, sino la proximidad de los mundos en los que M simplemente desaparece o no causa B, pero todo lo demás se mantiene igual. Y (2') se cumple por la razón simétrica. Pero, como digo, para mantener esto necesitamos que nos sean más cercanos aquellos mundos en los que si F no se da, o no causa B, B se da igualmente sin que la anteceda ninguna F* que aquellos en los que B es antecedido causalmente por alguna F*. Sin embargo, es muy dudoso que esto sea así. Quizás el principio del cierre causal del mundo físico no sea metafísicamente necesario, aunque creo que se podría defender su necesidad a posteriori (véase Teller (1984), pero desde luego, es muy razonable suponer que los mundos más cercanos están sometidos a nuestro principio, tal y como lo exponíamos al comienzo: "[n]ada puede afectar la distribución de materia y energía en el espacio-tiempo salvo la instanciación de propiedades básicas en los objetos básicos que ocupan el espacio-tiempo".

En resumen: El problema de la exclusión explicativa no puede ser solucionado con el recurso a la sobredeterminación. En primer lugar, existe el problema de las coincidencias masivas, que sólo puede evitarse estableciendo una relación entre mente y cuerpo, esto es, negando la sobredeterminación. Después, no parece haber ninguna forma razonable de mantener la verdad del enunciado de sobredeterminación: los contrafácticos (1') y (2') no pueden ser ambos verdaderos. Quizás podrían serlo si el principio del cierre causal del mundo físico no tuviera fuerza modal alguna, pero tal cosa es harto implausible¹.

Notas

¹ Agradezco al Gobierno Vasco la beca concedida para investigar en estos temas. Mis agradecimientos van también para: Agustín Arrieta, Stuart Silvers, Manuel García-Carpintero, Manuel Liz y Alfonso García Suárez (tribunal de mi tesis), y Jesús Ezquerro (director). Leí una versión de este artículo en el II Congreso de la SEFA, y otra, más breve, en el ESPP-98: agradezco los comentarios que me hicieron Josep Corbí y Gene Witmer, respectivamente, así como los de los informantes anónimos de esta revista. Gracias finalmente a Antonio Blanco Salgueiro, a quien, al escribir esto, quise sacar del error.

BIBLIOGRAFIA

- Block, Ned: 1990, 'Can the Mind Change the World?', in Boolos (ed.): *Meaning and Method: Essays in Honor of Hilary Putnam*, Cambridge, Cambridge University Press.
- Burge, Tyler: 1993, 'Mind-Body Causation and Explanatory Practice', in Heil y Mele (eds.): *Mental Causation*, Oxford, Clarendon Press.
- Crane, Tim: 1995, 'Mental Causation', *Proceedings of the Aristotelian Society* 69, 211-36.
- Fodor, Jerry: 1987, *Psychosemantics*, Cambridge, Mass., MIT Press.
- Horgan, Terence: 1989, 'Mental Quausation', *Philosophical Perspectives* 3, 47-76.
- Horgan, Terence: 1993, 'From Supervenience to Superdupervenience: Meeting the Demands of a Maternal World', *Mind* 102, 555-86.
- Horgan, Terence: 1997, 'Kim on Mental Causation and Causal Exclusion', *Philosophical Perspectives* 11, 164-181.
- Kim, Jaegwon: 1993, 'Epiphenomenal and Supervenient Causation', in *Supervenience and Mind*, Cambridge, Cambridge University Press.
- Kim, Jaegwon: 1998, *Mind in a Physical World*, Cambridge, Mass., MIT Press.
- LePore y Loewer: 1987, 'Mind Matters', *Journal of Philosophy* 93, 630-642.
- Poland, Jeffrey: 1994, *Physicalism*, Oxford, Clarendon Press.
- Sturgeon, Scott: 1998, 'Physicalism and Overdetermination', *Mind* 107, 411-433.
- Van Gulick, Robert: 1993, 'Who's in Charge Here? And Who's Doing all the Work?', in Heil y Mele (eds.): *Mental Causation*, Oxford, Clarendon Press.
- Teller, Paul: 1984, 'A Poor's Man Guide to Supervenience and Determination', *The Southern Journal of Philosophy* XXII, 137-163.

Agustín Vicente ha cursado sus estudios de Filosofía en Deusto, Londres y Barcelona, donde se graduó, y en la Universidad del País Vasco, donde ha llevado a cabo el doctorado. Es reciente doctor en Filosofía -Diciembre de 1998- con la tesis 'La teoría representacional de la mente y la exclusión causal-explicativa', y sus áreas de interés son la Filosofía de la Mente, Filosofía del Lenguaje y Metafísica.