Teorema
Vol. XVIII/1, 1999, pp. 155-70

Searle y el problema de la exclusión causal. Vindicación del materialismo frente al naturalismo biológico*

David Pineda

ABSTRACT

In this paper I offer two reasons which favour usual materialist accounts of the mind in front of Biological Naturalism as a solution to the mind-body problem. The first one says that Searle's metaphysics gets inescapably trapped by the Causal Exclusion problem, whereas it can be shown, on my view, that certain reductivist and functionalist theories of the mind may deal promisingly well with this problem. The second one says that Searle's main argument against Materialism, the explanatory gap, renders his metaphysics ill-founded, and so that it cannot be claimed as a reason for Biological Naturalism.

RESUMEN

En este artículo ofrezco dos razones para preferir las teorías materialistas usuales al naturalismo biológico como solución al problema mente-cuerpo. La primera razón es que la metafísica de la mente de Searle se ve irremediablemente atrapada por el problema de la exclusión causal, mientras que a mi juicio es posible desarrollar teorías reductivistas y funcionalistas que ofrecen buenas perspectivas con respecto a este problema. La segunda razón es que el principal argumento de Searle contra las diversas formas de materialismo, el problema del hiato explicativo, hace que su metafísica de la mente sea completamente infundada, y por tanto no constituye argumento alguno en su favor.

I. INTRODUCCIÓN

Al principio de su libro, *The Rediscovery of the Mind*, John Searle asegura que el problema mente-cuerpo tiene una solución simple. La solución consiste en reconocer que los estados mentales conscientes son un producto causal de ciertos estados y procesos neurológicos que tienen lugar en el cerebro humano y en el de otros animales superiores. La conciencia es un fenómeno biológico propio de ciertos organismos, como puedan serlo la digestión o la mitosis, aunque con la crucial diferencia de que la conciencia es ontológicamente subjetiva. Es decir, no se trata meramente de que tengamos un acceso epistémico subjetivo a nuestros estados conscientes, un tipo de acceso característico que no tenemos en el caso de la digestión o la mitosis, sino de que

nuestros estados conscientes existen de un modo subjetivo [Searle (1992), p. 94].

Searle bautiza su posición como "naturalismo biológico" y la confronta con las soluciones materialistas y dualistas usuales del problema mente-cuerpo. A estas últimas las rechaza todas con el argumento de que se basan en un distinción entre lo mental y lo físico que es, según él, conceptualmente confusa y metafísicamente incorrecta, y que impide reconocer el hecho de que los estados conscientes existen de un modo subjetivo. Ello lleva en todos los casos a sostener posiciones absurdas sobre la mente, por ejemplo a dudar de la existencia de los estados conscientes o de su eficacia causal en la acción.

Más adelante discutiré la noción clave (y a mi juicio oscura) de la metafísica de Searle de la existencia subjetiva de los estados conscientes, y por qué no lleva a ninguna solución aceptable del problema mente-cuerpo. Pero antes quisiera mostrar que de la metafísica de Searle, a pesar de que él piense lo contrario, se siguen el mismo tipo de dudas sobre la eficacia causal de la mente que asolan a ciertas teorías materialistas antirreductivistas, dudas provocadas por el argumento de la exclusión causal [Kim (1989b), (1992) y (1993a)], sólo que en el caso de la metafísica de Searle las perspectivas de despejar estas dudas son menos halagüeñas que en el caso de estas teorías materialistas. Discutiré también lo que considero el principal argumento de Searle contra el materialismo, que no es otro que el problema del "hiato explicativo" ("*explanatory gap*"), y argumentaré que la solución de Searle, que consiste esencialmente en aceptar este fenómeno, conduce a una metafísica insostenible por carecer de fundamento. El resultado de la discusión será doble. Por un lado el problema del hiato explicativo es ciertamente una dificultad que el materialismo todavía no ha podido resolver, eso hay que concedérselo a Searle, pero asumirlo sin más como hace él no lleva a ninguna solución aceptable del problema mente-cuerpo. Por otro lado, el materialismo está en mejores condiciones que el naturalismo biológico a la hora de afrontar el problema de la exclusión causal.

II. Dos argumentos contra el naturalismo biológico

Como reza el título de esta sección, en ella pretendo ofrecer dos argumentos contra la metafísica que Searle propone como solución al problema mente-cuerpo. El primero consiste esencialmente en una versión del argumento de la exclusión causal que muestra que se sigue de los principios metafísicos propuestos por Searle que los estados mentales conscientes no pueden ser causalmente eficaces en cuanto tales, es decir, en virtud de ejemplificar propiedades mentales. El segundo consiste en señalar una inconsistencia en la posición de Searle que yo al menos no sé cómo resolver.

Comencemos por el primer argumento. La metafísica de la mente que propone Searle se articula entre otros en los siguientes dos principios:

> *Principio de superveniencia causal*: estados neurofisiológicos del mismo tipo tienen como efecto causal estados mentales del mismo tipo. Así, si dos cerebros fueran neurofisiológicamente indiscernibles, entonces serían también mentalmente indiscernibles, aunque la relación modal inversa no tiene por qué darse [Searle (1992), pp. 124-5].

> *Principio de reducción causal*: los poderes causales de un estado mental se explican enteramente a partir de los poderes causales de los estados neurofisiológicos sobre los que sobreviene causalmente [ibíd., pp. 114 y 116].

La combinación de ambos principios da lugar a una metafísica con bastantes puntos de contacto con las teorías materialistas usuales. El principio de reducción causal genera una relación de dependencia entre el nivel mental y el nivel neurofisiológico, que consiste en parte en una relación de superveniencia. Lo extraño (con respecto, al menos, a las versiones usuales del materialismo) es que esa relación de superveniencia sea causal y no constitutiva, es decir, una relación modal entre propiedades o tipos de estados que no es causal [véase Kim (1984a)]. Sin embargo, según Searle, que los estados neurofisiológicos causan estados mentales es una obviedad que se desprende de lo que ya sabemos sobre la fisiología del cerebro [Searle (1992), p. 125][1].

Sin entrar a discutir ahora estos principios, lo que me propongo mostrar de momento es que conducen al escepticismo con respecto a la eficacia causal de los estados mentales en cuanto tales, o sea, en virtud de su ejemplificación de propiedades mentales. Veamos por qué.

Searle quiere reconocer tanto casos en que un estado mental *M* es causalmente eficaz con respecto a otro estado mental *N*, como casos en los que un estado mental *M* es causalmente eficaz con respecto a un estado físico, un movimiento corporal, *C*. Pero, como veremos, ninguno de estos casos parece compatible con los dos principios metafísicos apuntados.

Comencemos por el primer caso. Se sigue del principio de superveniencia causal que existen estados neurofisiológicos *P* y *Q* tales que *P* causa a *M* y *Q* causa a *N*. Es absurdo suponer que *M* causa a *N* y *P* no causa a *Q*, pues esto va contra el principio de reducción causal. Si ocurriera un caso así, entonces no podríamos explicar los poderes causales de *M*, en particular su poder para causar la ejemplificación de *N*, en términos de los poderes causales de *P*. Por tanto, si queremos que *M* cause a *N* debemos admitir que *P* causa *Q*. Hay que decir, además, que Searle reconoce abiertamente que esto debe ser así [Searle (1995), p. 119]. Ahora bien, si *P* causa a *Q* tenemos, por transitividad de la relación causal, que la ejemplificación de *M* y la ejemplifica-

ción de *N* son un doble efecto de la ejemplificación de *P*. Por tanto, no hay ninguna razón para entender que la relación modal entre *M* y *N* sea causal, de hecho hay todas las razones para pensar todo lo contrario, pues se trata de efectos de una causa común.

Searle podría objetar, tal vez, que *M* es, como *P*, causalmente eficaz con respecto a la ejemplificación de *Q*. Es decir, la idea sería que *Q* está sobredeterminada por *P* y por *M*. Esta posibilidad da lugar a una situación que Kim ha llamado de "causalidad descendente" ("*downward causation*"), en la cual un estado superviniente y el estado sobre el que subviene coinciden en un efecto causal. El argumento de Kim para rechazar esta situación apela esencialmente al principio de clausura causal de la física, según el cual causas físicas son suficientes para efectos físicos[2]. Así, supuesto que *P*, o bien otro estado físico del que dependa *P*, es causalmente suficiente para *Q*, la pretendida eficacia causal de *M* con respecto a *Q* resulta cuando menos sospechosa y en el mejor de los casos redundante, resultados ambos contrarios a los propósitos de Searle[3].

De todos modos creo que los resultados de este argumento pueden reforzarse yendo más allá de las conclusiones del propio Kim. Puede mostrarse que una situación de causalidad descendente es incompatible con los principios metafísicos de Searle. La razón es simple. De acuerdo con el principio de reducción causal, los poderes causales de *M* se explican enteramente en términos de los poderes causales de *P*. Nótese, sin embargo, que una parte de los poderes causales de *P* es su poder para producir *Q*. Ahora bien, si la ejemplificación de *M* causa la ejemplificación de *Q*, entonces podemos explicar el poder causal de *P* para producir *Q* diciendo que la ejemplificación de *P* subviene a la ejemplificación de *M* y la ejemplificación de *M* causa la ejemplificación de *Q*. Por otro lado, puesto que por hipótesis *P* es causalmente eficaz con respecto a *Q*, podemos explicar el poder causal de *M* para producir *Q* diciendo que la ejemplificación de *P* subviene a la ejemplificación de *M* y causa la ejemplificación de *Q*. El problema con estas dos explicaciones es que si una de ellas es correcta la otra tiene que ser correcta también, y tal cosa es imposible pues la relación de explicación es estrictamente asimétrica. Así, no puede ser que dispongamos de una explicación de (parte) de los poderes casuales de *M* en términos de los poderes causales de *P* y a la vez de una explicación de (parte) de los poderes causales de P en términos de los poderes causales de *M*. Deberíamos, pues, rechazar ambos tipos de explicaciones, pero el principio de reducción causal requiere que haya una explicación del primer tipo, y por otro lado ambas explicaciones parecen en sí mismas perfectamente correctas. Esta situación conceptualmente inconsistente se resuelve con toda facilidad abandonando el supuesto de que la ejemplificación de *M* sea causalmente eficaz para la ejemplificación de *Q*. Nótese que no apelamos aquí para nada al principio de clausura causal y mucho menos a un principio de exclusión explicativa. Mi versión del argumento es sim-

plemente que la idea misma de causalidad descendente es inconsistente con la metafísica de Searle.

El otro caso concierne a la eficacia causal de un estado mental *M* con respecto a un estado físico, un movimiento corporal, *C*. Searle entiende una acción como un compuesto de lo que él llama una "intención en la acción" ("*intention in action*") y un movimiento corporal, pero aún así considera que la ejecución de una acción conlleva siempre un tránsito causal de un estado mental a un estado físico [Searle (1983), capítulo 3]. Pues bien, de acuerdo con el principio de superveniencia causal, en una situación así debe haber un estado físico *P* cuya ejemplificación cause la ejemplificación de *M*. En caso de que quisiéramos sostener que *M* causa *C* tendríamos, por transitividad de la relación causal, nuevamente una situación de causalidad descendente, que nos llevaría, por las razones esgrimidas anteriormente, a rechazar el supuesto de causalidad mental-física.

Así pues, nuestra conclusión hasta ahora es que se sigue de los dos principios metafísicos que propone Searle que un estado mental no es nunca causalmente eficaz ni con respecto a otro estado mental ni tampoco con respecto a un estado físico. Si la relación entre los estados mentales y los estados físicos es la que nos propone Searle entonces los primeros son epifenómenos de los segundos, en contra de lo que el propio Searle pretende.

Éste es pues mi primer argumento contra el naturalismo biológico. Mi segundo argumento consiste en mostrar que el primer principio metafísico de Searle resulta completamente infundado a la luz del tercer principio, que aún no hemos comentado, según el cual los estados mentales existen de un modo subjetivo. Como ya indiqué antes la noción de subjetividad ontológica que maneja Searle, por muy obvia que pueda parecerle a él, no me parece clara en absoluto. Searle trata de hacerla comprensible repitiendo en todo el libro [Searle (1992)] afirmaciones como las que aparecen en el siguiente texto:

> ¿Qué más puede decirse acerca del modo subjetivo de existencia? Bien, en primer lugar es esencial darse cuenta de que en razón de su subjetividad, el dolor no es igualmente accesible a cualquier observador. Su existencia, podríamos decir, es una existencia de primera persona […] la ontología de lo mental es irreduciblemente de primera persona [Searle (1992), pp. 94-5][4].

La cuestión es qué quiere decirse con que la ontología de lo mental es irreductiblemente de primera persona. En la medida en que entiendo lo que quiere decirse creo que implica que los estados mentales conscientes sólo son epistémicamente accesibles a quien los padece. Esto es, no es meramente que cada ser humano tenga un modo característico, no compartido con otros seres humanos, de acceder a sus propios estados mentales, sino que ese modo característico es el único modo de acceder a ellos. Searle no aceptaría esta ca-

racterización por considerarla confusa. Según él, en el caso de un estado consciente no tiene sentido distinguir entre aquello a lo que accedemos y el acto mismo de acceder a ello, porque el estado consciente consiste precisamente en el acto de acceso [Searle (1992), pp. 97 y ss.]. Ahora bien, si mi estado consciente consiste en mi acto de acceso, si esa es su esencia, entonces es claro que no puede ser accesible a nadie que no sea yo.

Si esta consecuencia del carácter ontológicamente subjetivo de la consciencia es correcta (y no veo razón para pensar que no lo sea), entonces el principio de superveniencia causal, la tesis de que los estados neurofisiológicos causan estados mentales, deviene gratuito: no hay la menor razón para pensar que es correcto. Ello es así por lo siguiente. Una tesis causal es empírica; debemos, pues, en caso de creer, como cree Searle, que es correcta, poder señalar datos empíricos que la confirmen. Ahora bien, ¿qué tipos de datos empíricos son esos? De la ontología irreduciblemente de primera persona de lo mental y de la discusión de Searle sobre el problema de las otras mentes [Searle (1992), pp. 71-7], puede inferirse que se trataría de datos empíricos inicialmente en favor de la tesis de que mis estados neurológicos causan mis estados mentales, que confirmarían la tesis causal para mi caso, y a partir de ahí generalizaríamos a otros seres con cerebros semejantes a los nuestros con ayuda del principio "causas similares-efectos similares". Ciertamente, dado el carácter ontológicamente subjetivo de la consciencia, no se ve qué otro tipo de datos empíricos podríamos ofrecer.

Sin embargo, en realidad no podemos disponer de datos empíricos en favor de que mis estados neurológicos causan mis estados mentales. Si introduzco en mi cerebro un escáner sofisticado (sin duda, mucho más que los aparatos de los que dispone el neurólogo de hoy en día) y anoto pacientemente los resultados todo lo más que podré obtener son correlaciones entre ciertos estados mentales subjetivos (aquellos en los que consiste mi observación de los resultados del escáner) y ciertos otros estados mentales subjetivos (aquellos otros que padezco al hacer esas observaciones). Pero, dado el carácter ontológicamente subjetivo de la consciencia, no hay manera de inferir de esas correlaciones que estados neurológicos que corresponden al primer tipo de estados mentales (las observaciones) causan también el segundo tipo de estados mentales. Veámoslo con un ejemplo. Supongamos que cada vez que estoy en un estado observacional *O* padezco otro estado mental *M* (por ejemplo, siento dolor en la oreja izquierda). Para poder inferir de esto que cierto estado neurológico *N* es responsable de mi observación *O* del escáner y es causa de *M*, necesito disponer de una explicación de un estado subjetivo como *O* en términos neurológicos. Y esto es justamente lo que según Searle no puedo tener en ningún caso, pues si dispusiera de una explicación así entonces la consciencia no sería ontológicamente subjetiva al ser reducible a entidades objetivas, como las que postulan nuestras teorías neurológicas [Searle (1992), pp. 49 y 95]. Por

[Searle (1992), pp. 49 y 95]. Por otro lado, conjeturar que cierto estado neurológico *N* es la causa de *O* nos deja donde estábamos, pues de nuevo nos preguntaríamos que tipo de datos empíricos avalarían esta hipótesis causal, y el razonamiento nos llevaría a un regresión infinita inaceptable.

En suma, pues, el carácter ontológicamente subjetivo de los estados conscientes impide que haya el menor indicio empírico de una relación causal entre lo neurológico y lo mental y, por consiguiente, el primer principio metafísico de Searle parece infundado.

III. Por qué el argumento de Searle contra el materialismo no es un argumento a favor del naturalismo biológico

En sus críticas al materialismo [Searle (1992), capítulos 2 y 3], Searle distingue entre el materialismo reductivo, según el cual las propiedades mentales son propiedades físicas, y el funcionalismo, para el que las propiedades mentales serían propiedades funcionales múltiplemente realizables por propiedades físicas. En realidad ofrece dos argumentos contra el materialismo reductivo, pero el primero de ellos, la objeción de la múltiple realizabilidad de lo mental a través de lo físico, no me parece decisivo. La múltiple realizabilidad de un tipo de propiedad a través de otras puede deberse siempre a un fenómeno de ignorancia y no de heterogeneidad metafísica. Por ejemplo, dos conjuntos de partículas pueden parecer mecánico-estadísticamente muy distintos hasta que uno cae en la cuenta de que ejemplifican la misma energía cinética media. Puede incluso darse el caso de que no haya dos partículas en los dos conjuntos que ejemplifiquen la misma energía cinética. Así, un caso que hoy nos parece claro de múltiple realizabilidad, como el de cierto tipo de propiedades mentales, mañana puede ser un caso claro de reducción si descubrimos la propiedad física adecuada. Por tanto, alguien que por las razones que sea quiera aferrarse a una metafísica reductivista no dará su brazo a torcer solamente en función de que hay casos aparentemente claros de múltiple realizabilidad.

El segundo argumento puede ser resumido del siguiente modo. Cualquier identificación de una propiedad mental con una propiedad física va a ser una verdad *a posteriori*. Que la verdad sea *a posteriori* sólo puede explicarse porque la misma propiedad ha sido identificada, a través de conceptos diferentes, mediante propiedades distintas. Ahora bien, si las propiedades implicadas en el concepto mental son subjetivas, ello nos conduce al dualismo de propiedades; pero si no lo son, entonces la identificación deja fuera lo característicamente mental, la subjetividad, y por tanto no resuelve el problema mente-cuerpo [Searle (1992), pp. 36-7].

En mi opinión este argumento de Searle resulta muy confuso, pues en caso de que las propiedades implicadas en los conceptos mentales sean subjetivas, entonces no va a ser posible identificar a las propiedades mentales con propiedades físicas, a causa del problema del hiato explicativo. Es posible sin embargo que Searle tenga en mente un argumento similar desarrollado por Stephen White, según el cual el materialismo reductivo es falso a menos que el funcionalismo analítico —la tesis de que los conceptos mentales son conceptos funcionales— sea correcto [White (1986), sección 3]. La idea del argumento es la siguiente. Para que una identificación entre propiedades sea *a posteriori*, como es el caso de las propiedades mentales y las propiedades físicas, la misma propiedad debe de ser denotada por conceptos distintos de tal modo que no sea *a priori* cierto que ambos conceptos correfieren. Tomemos un ejemplo. Supongamos que queremos identificar el dolor con la estimulación de las fibras *C*. El único modo de explicar esta identificación *a posteriori* es suponer que el concepto de dolor y el concepto de estimulación de las fibras *C* denotan una misma propiedad mediante otras propiedades que esta propiedad tiene o con las cuales está relacionada. La cuestión es, qué tipo de propiedades pueden ser las propiedades implicadas en el caso del concepto de dolor. No pueden ser propiedades físicas, pues entonces sería a priori que el dolor tiene esas propiedades físicas o está relacionado con ellas, y no parece haber ningún enunciado de estas características que sea *a priori*. Por otro lado, si esas propiedades son mentales, nos amenaza una regresión infinita inaceptable. Por tanto, concluye White, las propiedades tienen que ser "atópicas" ("*topic-neutral*"), es decir, propiedades funcionales. Propiedades que una propiedad ejemplifica cuando capacita al estado en que se ejemplifica para tener un cierto rol causal. Son propiedades atópicas, pues en principio tanto propiedades mentales como físicas podrían ejemplificar propiedades funcionales[5].

En este punto podemos enlazar con las críticas de Searle al funcionalismo, que son críticas a la idea de que nuestros conceptos mentales sean conceptos funcionales, por ejemplo, de que nuestro concepto de dolor sea el concepto de un estado que desempeña un cierto papel causal. Las críticas se concretan en conocidos experimentos mentales, como la habitación china, el problema de los "*qualia*" ausentes o el espectro invertido. La posibilidad de concebir este tipo de situaciones lleva a pensar que un concepto funcional no puede recoger el aspecto cualitativo y subjetivo característico de nuestras experiencias de dolor.

Es cierto que todos estos experimentos mentales son discutibles. Por ejemplo creo que hay réplicas razonables al problema de la habitación china [Rey (1986)]. Y el trabajo de Shoemaker también ha mostrado discutible la concebibilidad de situaciones de "*qualia*" ausentes. Este autor defiende que lo característico de las experiencias mentales es que tengan cierto contenido cualitativo, no uno en concreto, y esto es algo que puede ser captado por un

concepto funcional [Shoemaker (1975) y (1981)]. Las ideas de Shoemaker no están tampoco libres de discusión [Block (1980) y White (1986)], pero arrojan dudas sobre la fuerza de estos experimentos mentales.

En cualquier caso no es oportuno entrar aquí en esta compleja y larga discusión. En mi opinión Searle lleva razón en que el Funcionalismo Analítico es en último término insatisfactorio, y la razón de ello, que ponen de manifiesto los distintos experimentos mentales, es el ya mencionado problema del hiato explicativo. Dicho muy brevemente el problema es que no parece haber modo de explicar el carácter subjetivo de nuestras experiencias mentales en términos intersubjetivos. Si bien la formulación del problema se remonta al famoso artículo de Thomas Nagel, Nagel (1974), tal vez la caracterización más vívida del mismo se encuentre en Jackson (1986). No hay más que imaginar a la pobre Mary, que sabe todo lo que hay que saber sobre la física del color y la fisiología de la visión humana, pero que encerrada en su laboratorio en blanco y negro desconoce el aspecto cualitativo de una sensación cromática de rojo. Sólo hay que preguntarse qué tipo de información, en una situación así, permitiría a Mary superar su desconocimiento que no sea la que adquiriría percibiendo cosas rojas, para sentir toda la fuerza del problema del hiato explicativo.

Searle lleva razón en que no se ha resuelto el problema del hiato explicativo y que tal problema constituye una objeción al materialismo (yo al menos desconozco que haya sido resuelto de un modo satisfactorio para el materialista). Sin embargo no puede utilizar este argumento contra el materialismo en favor de su naturalismo biológico. El naturalismo biológico consiste entre otras cosas en asumir que existe este hiato explicativo entre lo mental y lo físico (de ahí el carácter ontológicamente subjetivo de la consciencia). Pero como hemos visto al final de la sección anterior asumir tal cosa, lejos de suponer una solución del problema mente-cuerpo, lleva a una metafísica insostenible por infundada.

IV. El materialismo frente al problema de la exclusión causal

Hasta aquí hemos presentado dos argumentos contra el naturalismo biológico: su incapacidad para resolver el problema de la exclusión causal y evitar así el epifenomenismo mental, y el carácter infundado de su propuesta metafísica, que deriva de aceptar como un hecho el fenómeno del hiato explicativo. Frente a eso hemos visto que el materialismo tiene pendiente explicar dicho fenómeno, que ciertamente no puede asumir sin más. ¿Qué decir con respecto al problema de la exclusión causal? En esta última sección quiero dar razones para pensar que el materialismo ofrece mejores perspectivas de solución de este problema de las que ofrece el naturalismo biológico.

En primer lugar, en el caso del materialismo reductivo el problema simplemente no se plantea. Al ser, según esta teoría, las propiedades mentales propiedades físicas, no puede haber obviamente entre ellas ningún problema de compatibilidad causal. Sencillamente, se trata de las mismas propiedades.

Pero hoy en día, a causa del problema de la múltiple realizabilidad, la mayoría de los filósofos materialistas prefieren una versión no reductivista del materialismo, de acuerdo con la cual las propiedades mentales son múltiplemente realizables a través de propiedades físicas. Pues bien, lo que pretendo hacer en esta sección es proponer una teoría funcionalista de las propiedades múltiplemente realizables que parece libre del argumento de la exclusión causal de Kim, aunque al final mostraré también mis reservas con respecto a mi propia propuesta.

Mi teoría comienza suscribiendo una interpretación "de segundo orden" de las propiedades funcionales, según la cual un objeto o un estado *X* ejemplifica una propiedad funcional *F* si y sólo si *X* ejemplifica alguna propiedad que tiene un rol funcional determinado (el rol funcional asociado a *F*). Un rol funcional es una propiedad de segundo orden que ejemplifica otra propiedad *P* cuando *P* causa en aquello que se ejemplifica que mantenga ciertas relaciones causales con otros estados u objetos (aquellas que especifique el rol funcional) [Schiffer (1987), capítulo 2]. Así, la ejemplificación de una propiedad funcional *F* por parte de algo *X* exige siempre la ejemplificación por parte de *X* de otra propiedad —de ahí que hablemos de interpretación "de segundo orden" de las propiedades funcionales—, a saber, una propiedad que ejemplifique el rol funcional asociado a *F*. De este modo, por poner un ejemplo muy sencillo, podemos caracterizar la solubilidad como aquella propiedad funcional que ejemplifica un objeto si y sólo si ese objeto ejemplifica cualquier propiedad que causa que se disuelva al ser sumergido. Finalmente, un realizador de una propiedad funcional *F* es cualquier propiedad que ejemplifique el rol funcional asociado a *F*. Con lo cual si hay más de una propiedad así, *F* deviene múltiplemente realizable. Por ejemplo, la solubilidad tendrá tantos realizadores como propiedades físicas haya tales que su ejemplificación en un objeto cause la disolución de ese objeto en caso de inmersión[6].

Existe pues, según esta caracterización, una clara relación de dependencia entre una propiedad funcional y sus realizadores en el sentido de que sólo se ejemplifica la propiedad funcional en caso de que se ejemplifique un realizador suyo. Nótese, además, que la relación de dependencia es nómica, pues que una propiedad *P* sea un realizador de una propiedad funcional *F* depende de cuáles sean los poderes causales de *P* (qué otras propiedades causa o la causan), y eso es algo que no podemos establecer *a priori*[7]. Hasta aquí, pues, nuestra teoría funcionalista ofrece una buena base para una meta-

física materialista según la cual las propiedades múltiplemente realizables dependen nómicamente de sus realizadores físicos.

La cuestión a plantearse ahora es si es razonable pensar que las propiedades funcionales son causalmente eficaces. Éste es un problema amplio, en el espacio de que dispongo me ocuparé sólo de ofrecer una respuesta al problema de la exclusión causal.

En primer lugar, suponer que las propiedades funcionales son causalmente eficaces con respecto a propiedades que realizan a otras propiedades funcionales, o con respecto a propiedades físicas en general, nos comprometería con situaciones de causalidad descendente. Ahora bien, en mi opinión el materialismo debe de respetar una tesis parecida al principio de reducción causal que propone Searle, según la cual, los poderes causales de las propiedades múltiplemente realizables, y por tanto también de las propiedades funcionales, deben ser explicados enteramente en términos de sus realizadores físicos. De no ser esto así, habría leyes empíricas que no podrían explicarse a partir de leyes físicas, y esto me parece abiertamente incompatible con el materialismo. Pero si aceptamos esta tesis general, debemos considerar imposibles las situaciones de causalidad descendente, por las razones que dimos en la sección III.

Así pues, sólo parece tener sentido atribuir eficacia causal a una propiedad funcional con respecto a otras propiedades funcionales. El argumento de Kim contra esta posibilidad es en esencia que nos compromete de nuevo con un caso de causalidad descendente. Sean F y G dos propiedades funcionales y supongamos un caso en el que se ejemplifican ambas y queremos decir que la ejemplificación de F es causalmente eficaz con respecto a la ejemplificación de G. Sean ahora P y Q los realizadores de F y G respectivamente cuya ejemplificación ha hecho posible la ejemplificación de las propiedades funcionales. Claramente P causa Q (en caso contrario no podríamos pretender que F cause G). Pues bien, según Kim para que F cause G es necesario que cause también Q, pues la ejemplificación de G se produce gracias a la ejemplificación de su realizador Q. Es lo que él llama el principio de realización causal: cualquier causa de una ejemplificación de una propiedad múltiplemente realizable debe ser también causa de su realizador [Kim (1993a), pp. 205-6].

Es obvio que la aceptación de este principio nos compromete con la causalidad descendente y por tanto nos lleva a la conclusión de que las propiedades múltiplemente realizables no son causalmente eficaces. Pero ¿es necesario aceptar este principio? Uno podría sugerir que no razonando del siguiente modo. Existen relaciones causales entre propiedades funcionales sustentadas (o implementadas, como se dice a veces en la literatura) por relaciones causales entre sus realizadores, sin que haya relaciones causales entre las propiedades funcionales y sus realizadores. Así, por seguir con nuestro ejemplo esquemático, podemos decir que la eficacia causal de F con respecto a G se sustenta en la eficacia causal de P con respecto a Q. Pero la relación entre F y G, al ser ambas

múltiplemente realizables, es autónoma con respecto a la de sus realizadores, pues en cada caso es sustentada por relaciones causales entre realizadores distintos. De hecho, hace años el propio Kim propuso un modelo metafísico parecido [Kim (1984b)]. Sin embargo, ahora lo rechaza y su razón para rechazarlo es que un modelo así está comprometido con lo que llama el principio de herencia causal que dice que los poderes causales de una ejemplificación de una propiedad funcional son idénticos a los poderes causales de su realizador [Kim (1993a), p. 208]. De nuevo, aceptar este principio nos lleva al epifenomenismo de las propiedades funcionales. La razón de ello es clara. Al ser una propiedad funcional múltiplemente realizable tiene realizadores distintos con poderes causales distintos. De ser entonces cierto el principio de herencia causal, resultaría que ejemplificaciones distintas de la misma propiedad funcional tendrían poderes causales distintos, lo cual nos llevaría a identificarla con la disyunción de sus realizadores y a abandonar cualquier pretensión de eficacia causal para ella.

Kim desafía al materialista que quiera defender este modelo metafísico a caracterizar los poderes causales de las propiedades funcionales y múltiplemente realizables en general por un lado de un modo compatible con este modelo y por otro lado de un modo que no nos comprometa con el principio de herencia causal.

Pues bien, por mi parte considero que este reto es perfectamente asumible para el caso de las propiedades funcionales. La idea puede resumirse del siguiente modo. Nótese en primer lugar que se desprende de nuestra caracterización de segundo orden que aquello que caracteriza a una propiedad funcional es el rol funcional asociado a ella. Pues bien, mi propuesta consiste en caracterizar los poderes causales de una propiedad funcional en términos de las propiedades implicadas en su rol funcional, donde las propiedades implicadas en un rol funcional son aquellas propiedades implicadas por una descripción completa de ese rol funcional. Para comprender mejor esta noción clave, recurriré a un ejemplo algo más complejo, y no esquemático, de propiedad funcional.

Supongamos que el funcionamiento de una máquina expendedora de latas de coca-cola puede describirse correctamente mediante una teoría T que dice que la máquina consta de tres estados internos E_1, E_2, E_3, que verifican:

(1) Si se introducen en la máquina 100 pts. y la máquina está en E_1, entonces no emite ningún *output* y pasa al estado E_2.
(2) Si se introducen en la máquina 50 pts. y la máquina está en E_1, entonces no emite ningún *output* y pasa al estado E_3.
(3) Si se introducen en la máquina 100 pts. y la máquina está en E_2, entonces expende una coca-cola, devuelve 50 pts. y pasa al estado E_1.
(4) Si se introducen en la máquina 50 pts. y la máquina está en E_2, entonces expende una coca-cola y pasa al estado E_1.

(5) Si se introducen en la máquina 100 pts. y la máquina está en E_3, entonces expende una coca-cola y pasa al estado E_1.
(6) Si se introducen en la máquina 50 pts. y la máquina está en E_3, entonces no emite ningún *output* y pasa al estado E_2.

Podemos entonces entender estos estados internos como estados funcionales caracterizados por el rol causal que juegan en *T*, y definir la propiedad funcional P_1 como la propiedad que ejemplifica algo cuando ese algo tiene el rol causal que *T* especifica para E_1 en virtud de ejemplificar alguna otra propiedad (un realizador de P_1), y similarmente podemos definir las propiedades funcionales P_2 y P_3. Es importante, para evitar casos de causalidad descendente, entender que los estados *input* y *output* de la máquina pueden caracterizarse también funcionalmente, en función del rol causal que *T* determina para ellos, y que podemos definir consecuentemente las propiedades funcionales *input* y *output* correspondientes. Finalmente, parece intuitivamente claro que las propiedades funcionales definidas son múltiplemente realizables, pues puede haber máquinas físicamente muy distintas que satisfagan la teoría *T*.

Lo que propongo es individualizar los poderes causales de cada propiedad funcional, por ejemplo de P_1, en función de las propiedades implicadas en su rol funcional. Pero ¿cuáles son las propiedades funcionales implicadas en el rol funcional asociado a P_1? Por ejemplo, una de ellas es P_2. La razón es la siguiente: en la descripción completa del rol funcional asociado a P_1 se dice que en determinadas circunstancias el estado funcional E_1 debe causar el estado funcional E_2. Así, la descripción del rol funcional asociado a P_1 implica la propiedad de estar en el estado funcional E_2, es decir, la propiedad funcional P_2. Otro ejemplo. Si hemos caracterizado la propiedad *output* de expender una coca-cola como la propiedad funcional P_o, asociada al rol funcional que tiene este estado-*output* de acuerdo con *T*, entonces también podemos decir que P_o debe de aparecer en los poderes causales de P_2, pues *T* establece que en determinadas circunstancias E_2 debe de causar ese estado *output O*.

Nótese que bajo esta caracterización de los poderes causales de una propiedad funcional el principio de herencia causal es falso. Cualesquiera dos ejemplificaciones de una propiedad funcional tienen los mismos poderes causales, pues las propiedades implicadas en su rol funcional son las mismas, mientras que los poderes causales de sus realizadores son distintos. Pero además no hay casos aquí de causalidad descendente. Las propiedades que constituyen los poderes causales de una propiedad funcional son siempre propiedades funcionales, y las que constituyen los poderes causales de un realizador de una propiedad funcional son siempre realizadores de otras propiedades funcionales. Así, P_1 causa P_2, pero un realizador de P_1 sólo causa un realizador de P_2, según nuestra caracterización. Además, en caso de múltiple realizabilidad la relación causal entre P_1 y P_2 viene sustentada por

tiple realizabilidad la relación causal entre P_1 y P_2 viene sustentada por relaciones causales entre realizadores distintos.

Parece pues que hemos superado el reto de Kim. Nuestra caracterización de los poderes causales de una propiedad funcional es compatible con el modelo metafísico de causalidad sólo entre propiedades funcionales, evitando así los casos de causalidad descendente, y es incompatible con el principio de herencia causal.

¿Disponemos, pues, de una metafísica funcionalista capaz de responder al problema de la exclusión causal? Bien, tengo serias dudas al respecto. El problema está en que nuestra caracterización de los poderes causales de una propiedad funcional parece artificiosa. ¿Por qué tenemos que suponer que una propiedad funcional está relacionada causalmente con las propiedades implicadas en su rol funcional? Mas bien tenemos razones para pensar lo contrario. Una de ellas es un criterio epistémico de atribución de eficacia causal a una propiedad que juzgo correcto, a saber: si disponemos de una explicación completa de un fenómeno que no apela para nada a *P* como agente causal, entonces, *prima facie*, *P* no es causalmente eficaz para la producción de este fenómeno [Jackson y Pettit (1990), p. 198]. Ahora bien, en nuestro caso siempre tenemos una explicación causal de la ejemplificación de P_2 que no apela para nada a la fuerza causal de P_1: P_2 se ejemplifica al ejemplificarse un realizador suyo que ha sido, a su vez, causado por un realizador de P_1. La otra razón para ser escéptico es que una propiedad funcional mantiene relaciones conceptuales con las propiedades implicadas por su rol funcional, y eso hace *prima facie* implausible que mantenga también relaciones causales con ellas. Pero ésta es una última idea que no tengo espacio para desarrollar aquí.

Departament de Filología i Filosofía
Facultat de Lletres, Universitat de Girona
Pl. Ferrater Mora, 1, E-17071 Girona
E-mail: pineda@skywalker.udg.es

NOTAS

* Este trabajo ha sido posible gracias a la ayuda financiera del proyecto PB95-0760, subvencionado por la DGICYT. Agradezco a Manuel García-Carpintero, Terence Horgan y Joan Pagès sus comentarios a una versión previa de este trabajo que han contribuido a mejorarlo sustancialmente.

[1] Esta idea de Searle me parece fruto de una confusión, como lo es su idea de que la solidez de la mesa sobre la que escribo es producto causal de su estructura molecular. Sin embargo no dispongo aquí del espacio suficiente para argumentar este punto. De

cualquier modo veremos que la tesis de la superveniencia causal agrava considerablemente el problema de la Exclusión.

[2] Esta formulación del principio, usual en la literatura, presupone el determinismo. Sin embargo puede darse una versión que no lo presuponga y que no altera sustancialmente el argumento: para cada propiedad física *F* existe una explicación causal completa de cada ejemplificación de *F* que apela únicamente a propiedades físicas. Por comodidad, en el texto me atendré a la versión más simple.

[3] Kim dice en algunos escritos que la suficiencia causal de *P* con respecto a *Q* excluye la eficacia causal de *M* en virtud del principio de exclusión explicativa (por ejemplo en Kim (1989b), p. 281 y en Kim (1993a), p. 354), pero esto no parece ser así. El principio de exclusión explicativa dice que no puede haber dos explicaciones completas e independientes de un mismo acaecimiento, y aclara que una explicación que apele a una propiedad superviniente no es independiente de otra que apele a una propiedad subviniente [Kim (1989a), p. 251]. Por tanto, el principio no se aplica al caso que estamos examinando. Para que *P* excluya a *M* como causa hace falta un principio más fuerte que el que utiliza Kim que diga que si una propiedad es causalmente suficiente para un efecto entonces ninguna otra propiedad es causalmente eficaz para ese efecto. Este principio se utiliza, sin justificar, en la versión que Stephen Yablo da del argumento en Yablo (1992).

[4] Texto original: "What more can we say about this subjective mode of existence? Well, first it is essential to see that in consequence of its subjectivity, the pain is not equally accessible to any observer. Its existence, we might say, is a first-person existence […] the ontology of the mental is an irreducibly first-person ontology".

[5] El origen de este argumento está en Smart (1971).

[6] La interpretación "de primer orden" de las propiedades funcionales consiste en identificar sus ejemplificaciones con las ejemplificaciones de sus realizadores. Así, en caso de que la propiedad funcional sea múltiplemente realizable, ésta deviene disyuntiva. Según me parece, ésta es la posición que defiende David Lewis [Lewis (1980) y (1994)].

[7] Propongo aquí individualizar los poderes causales de una propiedad a partir de las otras propiedades con las que mantiene relaciones causales. Creo que esta caracterización recoge bien el sentido de esta noción implícito en la literatura, y por tanto me atendré a ella.

REFERENCIAS BIBLIOGRÁFICAS

BLOCK, N. (1980), "Are Absent Qualia Impossible?", *The Philosophical Review*, vol. 89, pp. 257-74.

JACKSON, F. (1986), "What Mary didn't know", *The Journal of Philosophy*, vol. 83, pp. 291-5.

JACKSON, F. y PETTIT, P. (1990), "Causation in the Philosophy of Mind", *Philosophy and Phenomenological Research*, vol. 50, suplemento, pp. 195-214.

KIM, J. (1984a), "Concepts of Supervenience", *Philosophy and Phenomenological Research*, vol. 45, pp. 153-76. Reeditado en Kim (1993b).

—— (1984b), "Epiphenomenal and Supervenient Causation", *Midwest Studies in Philosophy*, vol. 9, pp. 257-70. Reeditado en Kim (1993b).

—— (1989a), "Mechanism, Purpose and Explanatory Exclusion", en Tomberlin J. E. (ed.), *Philosophical Perspectives 3, The Philosophy of Mind and Action Theory*, Atascadero, California, Ridgeview Publishing Company, 1989, pp. 77-108. Reeditado en Kim (1993b).

—— (1989b), "The Myth of Nonreductive Materialism", *Proceedings and Addresses of the American Philosophical Association*, vol. 63, pp. 31-47. Reeditado en Kim (1993b).

—— (1992), "'Downward Causation' in Emergentism and Nonreductive Physicalism", en Beckermann, A., Flohr, H. y Kim, J (eds.), *Emergence or Reduction?*, Berlin, De Gruyter, 1992, pp. 119-38.

—— (1993a), "The Nonreductivist's Troubles with Mental Causation", en Heil, H. y Mele, A. (eds.), *Mental Causation*, Oxford, Oxford University Press, 1993, pp. 189-210. Reeditado en Kim (1993b).

—— (1993b), *Supervenience and Mind*, Cambridge, Cambridge University Press. (Las referencias de páginas de los artículos de Kim corresponden a la numeración de este volumen.)

LEWIS, D. (1980), "Mad pain and Martian Pain", en Lewis, D., *Philosophical Papers Volume I*, Oxford, Oxford University Press, 1983, pp. 122-32.

—— (1994), "Lewis, D.: Reduction of Mind" en Guttenplan, S. (ed.), *A Companion to the Philosophy of Mind*, Oxford, Basil Blackwell, 1994, pp. 412-31.

NAGEL, T. (1974), "What Is It Like to be a Bat?", *The Philosophical Review*, vol. 83, pp. 435-50.

REY, G. (1986), "What's Really Going On in Searle's 'Chinese Room'", *Philosophical Studies*, vol. 50, pp. 169-85.

SCHIFFER, S. (1987), Remnants of Meaning, Cambridge, Mass., The MIT Press.

SEARLE, J. (1983), *Intentionality*, Cambridge, Cambridge University Press.

—— (1992), *The Rediscovery of the Mind*, Cambrige, Mass., the MIT Press.

—— (1995), "Consciousness, the Brain and the Connection Principle: A Reply", *Philosophy and Phenomenological Research*, vol 55, pp. 217-20.

SHOEMAKER, S. (1975), "Functionalism and Qualia", *Philosophical Studies*, vol. 27, pp. 291-15.

—— (1981), "Absent Qualia Are Impossible ó A Reply to Block", *The Philosophical Review*, vol. 90, pp. 581-99.

SMART, J. J. C. (1971), "Sensations and Brain Processes", en Rosenthal, D. (ed.), *Materialism and the Mind-Body Problem*, Englewood Cliffs, Prentice-Hall, pp. 53-66.

WHITE, S. L. (1986), "Curse of the Qualia", *Synthese*, vol. 68, pp. 333-68.

YABLO, S. (1992), "Mental Causation", *The Philosophical Review*, vol. 101, pp. 245-80.