UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE HUMANIDADES Y CIENCIAS DE LA EDUCACIÓN

Departamento de Filosofía

Tesis Doctoral

CAUSACIÓN MENTAL Y EXPLICACIÓN PSICOLÓGICA EN EL DEBATE CONTEMPORÁNEO SOBRE EL MATERIALISMO NO REDUCCIONISTA

AUTOR: M. SC. GUSTAVO FERNÁNDEZ ACEVEDO

Director: Dr. Manuel Comesaña

Codirector: Dra. Alicia Gianella

Julio de 2003

INDICE

AGRADECIMIENTOS

Deseo agradecer a quienes han colaborado, de una u otra manera, en la preparación de esta tesis. Diana Pérez y Marcelo Sabatés atendieron mis consultas con excelente disposición; sus comentarios me fueron muy útiles para comprender ciertas cuestiones difíciles y dar forma a algunos pasajes. A mi codirectora, Alicia Gianella, quien siempre me ha brindado su estímulo y asesoramiento desde mis primeros avances en el tema hace ya varios años.

Una mención aparte merece mi director, Manuel Comesaña. Su generosidad para con mi trabajo ha sido invariable. Para quienes no provienen originariamente, como es mi caso, del campo de la filosofía, puede hacerse difícil en ocasiones despojarse de ciertos hábitos de pensamiento adquiridos en la formación en una disciplina científica y captar el carácter distintivo de la reflexión filosófica. De él aprendí más sobre este carácter en nuestras conversaciones informales a lo largo de los años que en muchos cursos y seminarios dictados por prestigiosos especialistas. Si esto es lo que distingue al verdadero maestro en esta disciplina, creo que él sin dudas lo es.

Por último (pero no menos importante), quiero agradecer a mi esposa Alejandra. Su estímulo y su apoyo han sido siempre constantes e incondicionales. Quien alguna vez haya emprendido una tarea difícil, apasionante y exigente, como es la realización de un trabajo de esta naturaleza, sabrá comprender la invaluable ayuda que esto puede representar.

INTRODUCCIÓN

Esta tesis aborda algunos de los aspectos centrales de uno de los problemas más importantes que han surgido en el ámbito de la filosofía de la mente y la filosofía de la psicología en los últimos años: el problema de la exclusión causal/explicativa de los sucesos mentales.

Como muchos otros problemas filosóficos, el marco general del problema de la causación mental, esto es, el debate sobre la relación mente-cerebro, ha cambiado notablemente en las últimas décadas. De un problema puramente filosófico, abordado desde la antigüedad con las características herramientas de esta disciplina, se ha transformado en un problema filosófico-científico. Los avances logrados en el conocimiento del cerebro y de los procesos mentales, como así también en las técnicas utilizadas para estudiarlos, generaron en algunos autores la confianza en que el avance de la ciencia permitiría resolver lo que hasta ese momento la filosofía no había logrado. Este punto de vista se ve en ocasiones expresado en el rechazo a la reflexión filosófica sobre la causación mental, y en el intento de delegar en la ciencia, particularmente en la ciencia cognitiva, la tarea de resolver el problema. Esta confianza, si bien razonablemente fundada, parece excesiva. Los aspectos filosóficos del problema no parecen poder ser abordados a través de metodologías empíricas; sencillamente, están más allá de cualquier evidencia disponible. Este comentario, lejos de sugerir preferencias por una filosofía autosuficiente, tiene la función de señalar lo que, a nuestro juicio, debe ser una relación de complementariedad y no de competencia entre ciencia y filosofía. La reflexión filosófica actual sobre lo mental no puede realizarse con independencia de los resultados de la investigación fáctica, pero sostener esto no implica afirmar que tal reflexión deba subordinarse completamente a la ciencia y limitarse a un rol auxiliar de sistematización u organización de los resultados logrados en esa investigación.

Una de las formas en que la reflexión filosófica contemporánea sobre lo mental ha mostrado el intento de adecuarse a la ciencia es el surgimiento de doctrinas que pretenden ajustarse a una concepción científica del mundo. En particular, esta concepción implica adoptar un punto de vista básicamente materialista de la realidad. Ahora bien, no cualquier forma de materialismo ha sido

considerada viable; la forma específica de materialismo que ha gozado de general aceptación en las últimas tres décadas es el materialismo no reduccionista. Una u otra forma de la doctrina materialista no reduccionista ha constituido en el ámbito de la filosofía de la mente el equivalente a la 'concepción heredada' en el ámbito de la filosofía de la ciencia. Su influencia ha sido notable, y lo sigue siendo, pese a los serios problemas que debe enfrentar.[1] Un problema especialmente grave es el que se produce al tratar de explicar la causación psicofísica. Este problema, como veremos, tiene varios aspectos; nos ocuparemos en especial de uno de ellos, que puede ser planteado a través de la siguiente pregunta: dado que todo suceso físico tiene una causa que también es física, ¿cómo es posible que posea, además, una causa mental?

Los aspectos de nuestra concepción del mundo que se ponen en juego al discutir la posibilidad de la causación mental no son menores. Al decir de Kim (1998), hay importantes razones para preservar la causación mental. En primer lugar, la posibilidad de la acción humana requiere que nuestros estados mentales tengan efectos causales en el mundo. Las acciones voluntarias requieren que nuestras intenciones, deseos y creencias causen movimientos apropiados de nuestros miembros, y que éstos a su vez causen una reorganización de los objetos que nos rodean. En segundo lugar, la posibilidad de conocimiento humano presupone la realidad de la causación mental: la percepción requiere que los objetos y sucesos físicos a nuestro alrededor causen nuestras percepciones y creencias; el razonamiento involucra la causación de una nueva creencia por una antigua; la memoria es un proceso causal complejo que involucra interacciones entre experiencias, su almacenamiento físico y su recuperación en forma de creencias. Por estas razones, observa, no resulta extraño que haya filósofos que no estén dispuestos a renunciar a la causación mental sin importar la fuerza de las presiones en sentido contrario. Todas estas consideraciones adquieren plena significación cuando se analizan, por ejemplo, las implicancias del epifenomenismo, esto es, la tesis que niega la eficacia causal de las propiedades mentales. El epifenomenismo subvierte de

[1] Pinker (1997), al describir los fundamentos de la psicología evolucionista, señala que este programa se asienta en dos pilares: la psicología cognitiva (en particular, la Teoría Computacional de la Mente) y la biología evolucionista. Respecto de la primera, señala que constituye uno de los grandes avances en la historia intelectual, que permite 'resolver el problema mente-cerebro'. Tanto entusiasmo se ve apenas justificado si consideramos las dificultades que este enfoque debe enfrentar.

manera tan radical nuestras imágenes de sentido común y científica del mundo que parece una consecuencia que debe evitarse a toda costa.

El problema de la presunta ineficacia causal de los estados mentales tiene consecuencias que van más allá del contexto propiamente filosófico. En particular, una consecuencia *prima facie* del problema es la probable irrelevancia de las explicaciones psicológicas: si lo mental es causalmente inerte, no puede jugar un rol distintivo en explicaciones causales de la conducta y de otros sucesos mentales. Esta consecuencia, a su vez, parece tener serias implicancias para una ciencia autónoma (aunque no aislada) de lo mental, como es concebida actualmente la psicología. Si los sucesos mentales carecieran de eficacia causal, se produciría una situación probablemente sin precedentes en la historia intelectual: la constitución de una ciencia que, mientras supone un dominio propio de entidades y propiedades, se funda sobre la base de un error fundamental. La noción de explicación psicológica, por lo tanto, parece ser uno de los vínculos principales entre los problemas ontológicos de la causación mental y la investigación fáctica.

El eje central de la tesis consistirá, entonces, en el examen de la relación del problema de la causación mental con las cuestiones epistemológicas vinculadas con las explicaciones psicológicas. Consideraremos que la explicación constituye el nexo que vincula la investigación fáctica en psicología con las cuestiones más puramente filosóficas referentes a la ontología de lo mental. Sin embargo, es conveniente hacer notar que no todas las estrategias que apelan al análisis de la explicación psicológica como forma de enfrentar el problema de la exclusión recurren a las explicaciones provistas por la psicología científica; algunas de ellas, como veremos, intentan resolver el problema mediante el examen de una u otra doctrina filosófica sobre la explicación científica.

Diversas preguntas reflejan el vínculo entre el problema ontológico de la causación mental y la explicación psicológica: ¿afectan los argumentos de la exclusión causal a las explicaciones psicológicas?; si lo hacen, ¿en qué medida?; ¿en qué forma debemos reinterpretar nuestras explicaciones causales mentalistas como consecuencia de esos argumentos?; ¿debemos subordinar la explicación psicológica a la ontología de lo mental?; ¿o, por el contrario, debemos dar prioridad a la manera en que efectivamente explicamos los fenómenos mentales? Trataremos de

proporcionar una respuesta a éstos y a otros interrogantes de manera consistente con algunos de los supuestos relativos a la interacción ciencia-filosofía que hemos mencionado.

A lo largo de este trabajo defenderemos las siguientes tesis:

1. El problema de la exclusión causal es un problema real y acuciante para las principales versiones del materialismo no reduccionista que desean conservar la realidad de lo mental y su eficacia causal, junto con la autonomía metodológica de la psicología. No se trata de un problema que admita soluciones sencillas o su simple disolución mediante estrategias poco costosas en términos filosóficos.
2. Tal problema plantea reales dificultades al menos a una parte importante de las explicaciones psicológicas tanto de sentido común cuanto científicas. Sin embargo, no hay razones de peso para admitir que toda explicación psicológica se ve afectada por los argumentos de la exclusión.
3. Las posiciones que, aceptando que los argumentos de la exclusión son sólidos, pretenden salvar a las explicaciones mentales de forma tal que mantengan cierta continuidad con las que actualmente son aceptadas en psicología, deben ser capaces de preservar ciertas características básicas que reconocemos en tales explicaciones. En particular, deberían ser capaces de constituir el fundamento para la acción racional en psicología.
4. La apelación a la prioridad explicativa por sobre el análisis ontológico (estrategia 'deflacionista') no es satisfactoria, ya que descansa en parte sobre supuestos discutibles y no argumentados referentes al éxito explicativo de la psicología.
5. Las estrategias de división del *explanandum* (conocidas como 'del *explanandum* dual' o 'de los dos *explananda*'), sólo parecen ser capaces de salvar una parte de la causación mental, esto es, la causación de sucesos mentales por otros sucesos mentales, al precio de renunciar a la causación de sucesos físicos por sucesos mentales y viceversa.
6. Las ganancias explicativas que presuntamente se logran al aceptar ciertas clases de relaciones entre lo mental y lo físico deben ser consideradas en relación con la plausibilidad de las posiciones que nos veamos llevados a adoptar en cuestiones ontológicas.

7. El principio de exclusión explicativa, que ha constituido un aspecto importante de algunos de los argumentos de la exclusión causal, no tiene, por sí mismo, la fuerza suficiente como para posibilitar la eliminación de las teorías psicológicas y su reemplazo por teorías neurofisiológicas. Este principio, además, requiere de ciertas reformulaciones para ajustarlo a la existencia de múltiples determinantes objetivos.

La tesis se estructurará en ocho capítulos. En el capítulo I se expondrán los antecedentes del problema y algunos de los principales argumentos que han conducido a dudar de la viabilidad de las doctrinas materialistas no reduccionistas. En el capítulo II se describirán algunas de las principales consecuencias del problema, junto con una clasificación de las estrategias que se han propuesto para lidiar con él; además, se examinará someramente su relación con la evidencia empírica, tanto en lo que respecta a su pertinencia para la solución del problema cuanto en lo que se refiere a sus potenciales efectos en ciertos programas exitosos de investigación fáctica. En el capítulo III examinaremos la incidencia de los argumentos de la exclusión en la explicación psicológica, señalando áreas de la investigación psicológica en las cuales su impacto es igual o mayor que en la que frecuentemente se menciona como más afectada por ellos (psicología cognitiva), y describiremos ciertas clasificaciones de las explicaciones psicológicas, analizándolas respecto de si pueden ser consideradas explicaciones causales. En el capítulo IV describiremos ciertas posiciones que, admitiendo la ineficacia causal de los estados mentales (ya sea que tales estados sean causados por estados cerebrales o determinados no causalmente por ellos), intentan proporcionar la base para la preservación de las explicaciones psicológicas. En el capítulo V examinaremos las estrategias que pueden caracterizarse como 'deflacionistas' con respecto al problema, centrándonos especialmente en aquellas que conceden prioridad a la explicación por sobre la metafísica para nuestra comprensión de la causación mental. En el capítulo VI describiremos algunas variantes recientes de la denominada 'estrategia del *explanandum* dual', y discutiremos sus vínculos con el problema ontológico principal. En el capítulo VII analizaremos una forma reciente de dualismo, que intenta reconsiderar las posibilidades de esta antigua doctrina de resolver el problema de la causación mental, a la vez que defiende la ganancia

explicativa que tal doctrina supuestamente implicaría. Por último, en el capítulo VIII discutiremos los fundamentos y alcances del principio de exclusión explicativa, analizando la posibilidad de que constituya el fundamento del reemplazo de la teorización psicológica por teorías provenientes de las neurociencias.

Un tratamiento riguroso del tema enfrenta dificultades no menores, comenzando por la ingente cantidad de bibliografía publicada al respecto en los últimos años. Tanto los problemas planteados cuanto las alternativas de solución, muchas de ellas notablemente sutiles y elaboradas, requieren un gran esfuerzo de análisis; sin embargo, no cabe esperar menos al enfrentar uno de los problemas filosófico-científicos más difíciles. No puede evitarse en ocasiones la identificación con Thomas Nagel cuando afirma:

> [La filosofía] es una materia muy difícil, y no constituye una excepción a la regla general el hecho de que los esfuerzos creativos rara vez tienen éxito. No siento que ocurre esto con los problemas que se tratan en este libro. Me parece que requieren un nivel de inteligencia por completo distinto del mío. Quienes hayan intentado abordar las preguntas centrales de la filosofía reconocerán esta sensación.[2]

[2] Thomas Nagel (1986), *Una visión de ningún lugar*, México, Fondo de Cultura Económica, 1996.

CAPÍTULO I: EL PROBLEMA DE LA EXCLUSIÓN CAUSAL-EXPLICATIVA

1. El debate sobre la causación mental y el materialismo en las últimas décadas

El objetivo del presente capítulo es el de presentar sistemáticamente uno de los problemas centrales de la causación mental: el denominado 'problema de la exclusión causal-explicativa' (denominación que tomamos de Kim). Pero antes de proceder a esta descripción convendrá situar los antecedentes y orígenes del problema.

Como suele recordarse en los escritos filosóficos, los problemas filosóficos no surgen en un vacío conceptual. Si bien las cuestiones suscitadas por la causación mental son antiguas, los problemas modernos no son los mismos que debió enfrentar Descartes. Ya no se trata, contemporáneamente, de explicar la interacción causal entre sustancias de naturaleza esencialmente diferente. Desterradas las doctrinas dualistas de sustancias y descartado el conductismo filosófico como alternativas plausibles, en la segunda mitad del siglo XX la atención se volcó decididamente a las diversas variantes del materialismo. Las primeras doctrinas materialistas contemporáneas, en la forma de la Teoría de la Identidad, parecieron satisfacer las exigencias de ajustar una teoría de la mente a la visión científica del mundo, básicamente fisicalista. Sin embargo, transcurrida aproximadamente una década del surgimiento de la Teoría de la Identidad, nuevas formas de materialismo comenzaron a predominar, relegando a aquella a un lugar secundario.

El origen del debate actual sobre la causación mental puede situarse en el surgimiento, a fines de la década del '60, de doctrinas que, pretendiendo preservar una visión materialista del mundo, mantenían a la vez un status privilegiado para las propiedades o sucesos mentales. Estas propiedades serían irreductibles a propiedades físicas, pese a lo cual poseerían poderes causales autónomos, y servirían para la explicación y predicción de la conducta. El auge de estas doctrinas, en el cual mucho han influido célebres ensayos de Putnam (1967) y Davidson (1970), coincidió con la pérdida de popularidad de la Teoría de la Identidad, hasta ese momento teóricamente dominante. Con el surgimiento de estas nuevas formas de

materialismo (fundamentalmente el monismo anómalo y el funcionalismo) el problema mente-cerebro y las cuestiones suscitadas por la causación mental parecieron quedar definitivamente resueltos. No obstante, los problemas reaparecerían más temprano que tarde en una nueva forma. En particular, los problemas de la causación mental resurgirían con notable fuerza, poniendo en riesgo la viabilidad misma de las doctrinas materialistas no reduccionistas.

Existen muchos ensayos recientes destinados a examinar el debate sobre la causación mental dentro de un marco fisicalista. Por mencionar sólo unos pocos: Beckermann (1992), Crane (1995), Hardcastle (1998), Heil (1991, 1992), Jackson (1996), Macdonald y Macdonald (1986). El espíritu general de los planteamientos de la cuestión gira básicamente en torno a lo siguiente: cómo reconciliar una visión fisicalista o materialista del mundo con la existencia de propiedades mentales no reductibles a lo físico pero que ejercen una influencia sobre él. Sin embargo, los problemas de la causación mental abordados en estos ensayos se vinculan con cuestiones bastante diversas: el funcionalismo, el externalismo, la identidad, el anomalismo, la eficacia causal del contenido, el reduccionismo, el emergentismo. Por esta razón, existen múltiples problemas de la causación mental y, correlativamente, muchas formas de plantear tales problemas.[3] Con el fin de situar el problema que nos ocupará y diferenciarlo de otros con los que se encuentra vinculado, convendrá en primer lugar distinguir los problemas básicos y los supuestos que los originan.

2. Tres problemas de la causación mental

Una manera sencilla y eficaz de situar el problema que nos interesa parece ser distinguir entre distintas cuestiones fundamentales relativas a la causación mental, identificando los supuestos a partir de los cuales surgen. Seguiremos en esto a Kim (1998). En este libro Kim distingue[4] tres problemas distintos de la causación mental.

[3] Pese a la diversidad de los problemas analizados en esos y otros estudios, parecería que algunas nociones básicas no han recibido la misma atención. Pérez (2000) señala que diversos autores han tomado a la noción de causalidad como 'básica e inanalizable', sin prestar atención al problema de proporcionar una explicación clara de lo que se entiende por causación o relación causal.
[4] Distinción que ya estaba presente en su (1991).

Cada uno de estos problemas está vinculado a distintas doctrinas con respecto a la mente.

La primera doctrina es la denominada 'del anomalismo de lo mental',[5] la cual sostiene la inexistencia de leyes psicológicas. El problema de la causación mental para esta doctrina puede ser formulado, en opinión de Kim, a través de la pregunta '¿cómo pueden ser propiedades causales las propiedades anómalas?' La segunda doctrina que origina problemas relativos a la causación mental es el computacionalismo y el externalismo con respecto al contenido. Para esta doctrina el problema puede ser formulado a través de la pregunta '¿cómo pueden ser propiedades causales las propiedades extrínsecas?' La tercera es denominada por Kim 'de la 'exclusión causal'; la pregunta correspondiente puede ser formulada de la siguiente forma: 'dado que todo suceso físico tiene una causa que es una causa física, ¿cómo es posible que haya también una causa mental?'

Cada una de estas doctrinas genera entonces distintos problemas de la causación mental que pueden ser tratados de manera independiente, si bien Kim considera que están en alguna medida interconectados y que una teoría comprehensiva de la causación mental debería proveer una solución satisfactoria a los tres. Los problemas originados en las dos primeras doctrinas, en su opinión, han sido más estudiados; el tercero es más reciente y se le ha prestado una menor atención. Pero no es ésta la diferencia más importante entre el último y los otros dos problemas. Según observa Kim, este problema, a diferencia de los restantes, golpea el corazón mismo del fisicalismo, ya que no contiene supuestos que no estén ligados esencialmente a esta doctrina. Por esta razón, además, puede tratarse independientemente de los restantes. En lo sucesivo nos ocuparemos únicamente de este problema. Cada uno de los demás problemas es de una complejidad suficiente como para merecer un análisis independiente.

En el planteamiento y consideración de la importancia del problema de la exclusión causal-explicativa han sido decisivos, sin ninguna duda, los análisis de Kim. En diversos ensayos (1989a, 1989b, 1990, 1991, 1993a, 1998), y ya a partir de fines de la década del '80, Kim ha señalado a este problema como la principal amenaza contra la plausibilidad de las doctrinas materialistas no reduccionistas. No

obstante, pese a que los análisis de Kim han sido sin duda de los más influyentes e incisivos, no han sido los primeros. Ya a fines de la década del '60 (esto es, paradójicamente en coincidencia con el inicio del período de auge de las doctrinas materialistas no reduccionistas) surgieron análisis tendientes a plantear –y, eventualmente, a resolver- la tensión existente entre explicaciones de la conducta que apelaran a estados intencionales y explicaciones que recurrieran a mecanismos causales físicos o neurofisiológicos. A algunos de estos análisis nos referiremos en el apartado que sigue.

3. Mecanismos y propósitos: los primeros planteos del problema

3.1. El planteo de Malcolm

El primer planteo explícito del problema de la exclusión (aunque sin que haya sido denominado de esta manera) puede situarse con cierta precisión en 1968, con el artículo de Norman Malcolm titulado 'The Conceivability of Mechanism'. En ese artículo, Malcolm planteaba como cuestión central la posibilidad de existencia de una teoría neurofisiológica 'suficientemente rica como para proveer explicaciones causales sistemáticas de todos los movimientos corporales no debidos a causas físicas externas' (p. 45). Esta teoría neurofisiológica sería, asimismo, apropiada para *predecir* todos los movimientos corporales. Una teoría tal, enfatizaba Malcolm, sería 'no propositiva' [*nonpurposive*], indicando con esto que no contendría conceptos tales como deseos, objetivos, metas, motivos o intenciones. En este sentido, contrastaría fuertemente con las explicaciones propositivas usuales de la conducta con las que estamos familiarizados, las cuales refieren la conducta a propósitos, deseos, metas e intenciones.

Las explicaciones provistas por tal teoría neurofisiológica diferirían de las explicaciones propositivas en tres aspectos relevantes. En primer lugar, pertenecerían a una teoría comprehensiva, mientras que las explicaciones propositivas no lo harían. En segundo término, como ya se dijo, no emplearían conceptos tales como propósitos o intenciones. Por último, incluirían leyes contingentes, mientras que las explicaciones propositivas no lo harían. Esta última

[5] Debida a Donald Davidson. *Cfr*. en especial su (1970).

diferencia, debido a su importancia para la posibilidad de reducción de una teoría a la otra, es desarrollada por Malcolm con cierto detalle.

Siguiendo el modelo nomológico-deductivo hempeliano, Malcolm sostiene que una de las proposiciones constituyentes del *explanandum*, en el caso de las explicaciones neurofisiológicas, sería del tipo 'Toda vez que un organismo de estructura *S* esté en un estado neurofisiológico *q* emitirá un movimiento *m*.' Por el contrario, la primera premisa de una explicación propositiva consistiría en una afirmación del tipo 'Toda vez que un organismo *O* tenga la meta *G* y crea que la conducta *B* es requerida para lograr *G*, *O* emitirá *B*.' Para Malcolm, el diferente *status* epistémico de las dos premisas se pone de manifiesto ante la situación de determinar en que circunstancias serían verdaderas. En ambos casos se hace necesaria la adición de cláusulas *ceteris paribus* (entendiendo por esto en este caso 'con tal de que no haya factores que contrarresten' [*countervailing factors*]). Supóngase entonces que se desea explicar por qué un hombre sube a un techo utilizando una escalera. Es el caso que el viento ha llevado su sombrero hasta el techo, y quiere recuperarlo. La primera premisa de una explicación propositiva, en este caso, sería la siguiente: 'Si un hombre quiere recuperar su sombrero y cree que esto requiere subir a una escalera, lo hará con tal de que no haya factores que contrarresten' (*ibid*., p. 48). Factores que contrarresten, en este caso, serían la carencia de una escalera, el miedo a la subida, etcétera. Pero en este caso, observa Malcolm, la adición de cláusulas *ceteris paribus* transforma a la primera premisa de la explicación en una proposición *a priori*. Si no existieran riesgos ni impedimentos, reales o imaginarios, de tipo físico o psicológico, entonces si el hombre no subiera la escalera no sería verdad que *quería* su sombrero, o que *se proponía* recuperarlo.[6]

Las consecuencias teóricas de esta diferencia en las premisas de ambos tipos de explicación son, para Malcolm, de máxima importancia. Dado que los principios de las explicaciones propositivas son 'principios de acción' *a priori*, no es posible que sean 'menos básicos' o 'dependientes' de los principios de las explicaciones neurofisiológicas, que constituyen leyes contingentes.[7] Una segunda consecuencia de

[6] Malcolm señala que esta diferencia entre las premisas de ambas explicaciones ha hecho pensar a algunos filósofos que las explicaciones propositivas no son explicaciones causales.

[7] Kim señala insuficiencias del concepto de dependencia que maneja Malcolm. En particular, advierte que 'la noción de 'dependencia' utilizada es demasiado estrecha y en el mejor de los casos parece caracterizar

importancia reside en el hecho de que la confirmación de una teoría neurofisiológica comprehensiva *no prueba* que los principios de las explicaciones propositivas sean falsos. Y no puede hacerlo ya que estos principios son verdaderos *a priori*: no pueden ser refutados.

Una vez que ha especificado las diferencias entre los *explanans* de ambos tipos de explicaciones, Malcolm dirige su atención directamente a la posibilidad de que éstas no sean realmente rivales. Esto es lo que han sostenido algunos filósofos, comenta, al afirmar que los *explananda* de ambas explicaciones no son equivalentes: mientras que la explicación neurofisiológica explica movimientos, la explicación propositiva explica acciones. Ya que ambos tipos de explicación emplean distintos conceptos y supuestos, sería posible que coexistieran, dado que cada una sería irrelevante para la otra.[8] Malcolm descarta esta posibilidad afirmando que las distintas explicaciones pueden compartir, como *explanandum*, la realización de una misma acción, por ejemplo, 'de una y la misma ocurrencia de la subida de un hombre por una escalera' (*ibid*., p. 52). En este caso, sostiene, se supone que la teoría neurofisiológica imaginada proveería de explicaciones causales *suficientes* de la conducta en términos neurofisiológicos; no habría lugar en ella para cosas tales como deseos o intenciones.

Una posible salida de la situación de exclusión entre ambas explicaciones la constituiría la alternativa de la identidad: si la condición neural del hombre que causa su subida por la escalera fuese contingentemente idéntica a la intención del hombre de subirla, no tendríamos dos explicaciones sino sólo una. Malcolm rechaza de manera tajante esta posibilidad, afirmando que la idea de identidades contingentes entre intenciones y condiciones neurales es una hipótesis no significativa. Intentar verificar esta hipótesis implicaría el tratar de descubrir en que parte del cerebro de *A* está situada una determinada intención, lo cual no tiene sentido.[9] Y si se pretendiera estipular que la localización de una determinada intención es la misma localización

---

a un subcaso especial; en segundo lugar, su argumentación hace uso de suposiciones especiales que necesitan justificación y explota lo que parecen ser rasgos locales del caso particular entre manos' (1989a, p. 240).

[8] La alternativa de considerar que las explicaciones neurofisiológicas y propositivas no comparten el mismo *explanandum* ha sido denominada estrategia del '*explanandum* dual'. Esta estrategia ha sido retomada posteriormente por otros autores, en especial por F. Drestke y por Ausonio Marras. Volveremos sobre esta cuestión en el capítulo VI.

[9] Esta es una de las conocidas objeciones a la teoría de la identidad de tipos.

de sus procesos neurales correlativos, la identidad que resultaría de esta correlación ya no sería contingente.

Malcolm concluye su exposición afirmando que el mecanicismo, en la forma en que lo presenta, *es concebible*, ya que nada en su examen indica que constituya una doctrina autocontradictoria; el universo podría estar poblado de organismos cuyos movimientos podrían ser completamente explicados en términos de la teoría neurofisiológica que imagina.[10] Ante la existencia de una teoría tal, no quedaría espacio para las explicaciones propositivas: ambos tipos de explicación, como hemos visto, no son compatibles.

3.2. La réplica de Goldman

En un artículo en respuesta al de Malcolm, Alvin Goldman (1969) sostuvo que las explicaciones propositivas y neurofisiológicas podrían ser compatibles. El objetivo de Goldman fue discutir y fundamentar el rechazo de la que consideraba la tesis central de Malcolm, esto es, la idea de que '*si* los estados neurofisiológicos son *suficientes* para la conducta, entonces los deseos o intenciones *no son necesarios* para la conducta' (p. 470. Cursivas del autor).

Goldman comienza su crítica rechazando el principio de suficiencia causal que sostiene la propuesta de Malcolm. Este principio, según él, sostiene que '(I) Si los sucesos C_1,…, C_h son conjuntamente suficientes para la ocurrencia del suceso E, entonces ningún suceso salvo C_1,…,C_h es necesario para la ocurrencia de E' (*ibid.*, p. 470). Sin embargo, pese a su plausibilidad inicial, prosigue, este principio excluye la posibilidad de una cadena de sucesos, tal que cada eslabón es una causa del eslabón subsiguiente, y todos ellos pueden ser considerados una causa del eslabón final de la

[10] Sin embargo, Malcolm admite que hay aspectos en los que el mecanicismo no es concebible. Dado que el mecanicismo es incompatible con la existencia de conducta intencional, y que el habla humana es, en su mayor parte, conducta intencional, se sigue que si el mecanicismo es verdadero, nadie podría declarar o afirmar nada. Habría entonces un absurdo lógico, que no implicaría que el mecanicismo fuese autocontradictorio, pero sí que impediría afirmarlo. Esto es, la proposición compuesta 'el mecanicismo es verdadero y alguien afirma que es verdadero', constituye una autocontradicción. Esto, no obstante, no constituiría una refutación del mecanicismo y no pondría en peligro su *status* de teoría científica. Si se confirmara una teoría que probara que las personas no tienen deseos, propósitos o metas, este resultado, dice Malcolm, tendría que ser aceptado sin importar lo perturbador que pudiera ser. Malcolm señala un segundo aspecto en el cual el mecanicismo sería no concebible. Hacer o decir algo por alguna razón implica que lo que se hace o dice constituye una conducta intencional. Pero dado que el mecanicismo es una doctrina que es incompatible con la existencia de intenciones, esto lleva a que las personas son incapaces de hacer o decir algo por una razón. Debido a esto, no podría haber algo tal como una razón

cadena. Dado que la exclusión de esta posibilidad le parece inaceptable, propone en su lugar un principio modificado de suficiencia causal: '(II) Si los sucesos C*, ocurriendo en t_1, son suficientes para la ocurrencia de E en t_2, entonces ningún otro suceso en t_1 es necesario para la ocurrencia de E en t_2' (*ibid.*, p. 473).[11]

Goldman considera que puede probarse que tanto el principio (I) cuanto el (II) son inaceptables, recurriendo para ello a la noción de 'equivalentes nómicos simultáneos'. Si ciertos sucesos (o conjuntos de sucesos) C* y C** son equivalentes nómicos simultáneos, esto significa que C* es suficiente para C** y C** es suficiente para C*; de manera equivalente, significa que C* es necesario para C** y C** es necesario para C*. Si ahora suponemos que C* es necesario y suficiente para la ocurrencia subsecuente de E en el momento t_2, entonces el otro suceso C**, que es simultáneo con C*, es también necesario para la ocurrencia de E. Esto se sigue por la transitividad de la relación 'A es necesario para B'. Dado que C** es necesario para C* (por su condición de equivalentes nómicos simultáneos) y C*, por hipótesis, es necesario para E, debe concluirse que C** es necesario para E. Si C* o C** no hubieran ocurrido, entonces E no habría tenido lugar. Este es un caso, según Goldman, en el cual el principio (II) resulta violado: hay un suceso causalmente suficiente para E que es acompañado por un suceso que es causalmente necesario para E.

Aplicando estos resultados al problema de la compatibilidad entre causación neurofisiológica y propositiva, Goldman considera que no hay razones para pensar que las intenciones o deseos no puedan ser equivalentes nómicos simultáneos con estados neurofisiológicos. Es lógicamente posible, y compatible con la información presente, sostiene, que debería haber leyes que establecieran, para cualquier organismo humano y para cualquier momento, que un individuo tendrá una intención específica en ese momento si y sólo si está en un estado neurofisiológico específico en ese momento. Considera que hay al menos dos maneras de defender la

---

para sostener que el mecanicismo es verdadero. Esta es una segunda paradoja, según Malcolm, del mecanicismo: éste no puede ser sostenido sobre bases racionales.

[11] Kim (1989a) considera que Malcolm no debería aceptar esta versión modificada de su principio, ya que, si lo hiciera, no podría mostrar la incompatibilidad entre una explicación propositiva y una neurofisiológica que hiciera uso de las condiciones fisiológicas iniciales que tienen lugar después o antes que la creencia o los deseos invocados en la explicación propositiva.

compatibilidad de causación neurofisiológica y causación propositiva: la teoría de la identidad y la hipótesis de los equivalentes nómicos simultáneos.

La posición que cree correcta, prosigue, no debe ser identificada con el epifenomenismo, el paralelismo o el interaccionismo. No es epifenomenista, sostiene, ya que no postula relaciones causales unidireccionales de sucesos físicos a sucesos mentales; debido a la simultaneidad de sucesos mentales y físicos, probablemente no debería llamarse 'causal' a la relación entre ellos. Por esta misma razón, resultaría inadecuado identificar a su posición con el interaccionismo, que postula la existencia de relaciones causales de estados mentales a estados físicos y viceversa. Por otra parte, debido a que postula relaciones nómicas entre ambos tipos de sucesos, tampoco debe considerarse a su posición como paralelista: la co-ocurrencia de ambas clases de sucesos no es meramente accidental, sino que hay una ley universal involucrada en la aparición de los equivalentes nómicos.

Tampoco debe considerarse que la solución de los equivalentes nómicos simultáneos implique la sobredeterminación de la conducta humana. El *status* de los equivalentes nómicos simultáneos difiere significativamente de casos de sobredeterminación como el de una persona que muere cuando le disparan dos tiradores simultáneamente. Este último caso, dice Goldman, difiere del primero en que involucra distintas entidades espacialmente distantes (los dos tiradores); en segundo lugar, los dos tiradores son independientes en el sentido de que ninguno es necesario ni suficiente para el otro, mientras que los equivalentes nómicos son tanto necesarios cuanto suficientes el uno para el otro.

3.3. Los primeros análisis de Kim

En uno de los primeros artículos en los que se ocupó del problema (1989a),[12] Kim comienza analizando la propuesta de Malcolm y la réplica de Goldman, utilizando este debate como marco inicial para su análisis. Su posición general con respecto a esta discusión es que el planteo de Malcolm es básicamente correcto, si bien esto no implica que Goldman esté equivocado. Esta aparente contradicción se desvanece cuando se analiza en detalle la respuesta que ambos autores darían al principio general que Kim extrae como conclusión de su análisis.

[12] Si bien hay referencias anteriores a la exclusión al menos en sus (1979) y (1987).

El 'principio de exclusión explicativa', que Kim enuncia, sostiene que a ningún acontecimiento puede dársele más que una explicación completa e independiente. La posición de Malcolm estaría fundamentalmente de acuerdo con este principio: las explicaciones neurofisiológicas y propositivas se excluyen mutuamente, dado que no son dependientes una de la otra (lo que Malcolm intenta probar con su afirmación relativa a las diferencias entre los respectivos *explanans*), y que ambas son completas. La respuesta de Goldman, si bien afirma que tanto la explicación neurofisiológica como la propositiva son legítimas y compatibles, no sostiene que ambas sean independientes: la noción de equivalentes nómicos simultáneos señala un nexo ontológico entre los respectivos *explanans*.

No obstante, Kim señala varios aspectos en los que la respuesta de Goldman puede ser discutida. En primer lugar, advierte sobre una insuficiencia de la tesis de los equivalentes nómicos simultáneos (a la cual, dice, podría denominarse 'tesis de la correlación psicofísica'). Esta tesis no excluye la posibilidad de que, en realidad, C** sea sólo un epifenómeno de C*, aunque sea nomológicamente suficiente (y quizás también necesario) para la ocurrencia de C*. Un ejemplo de esta posibilidad está dado por una situación tal que *A* es un estado patológico determinado por alguna enfermedad, *A** un síntoma simultáneo de ese estado y *E* un estado posterior de la enfermedad. En este caso, prosigue Kim, no hay dos explicaciones para un único *explanandum*, ya que el epifenómeno no explica.[13]

En segundo lugar, y dejando de lado la posibilidad de los epifenómenos, resulta dudoso de que existan leyes de correlación que permitan vincular estados intencionales (como creencias y deseos) con estados neurofisiológicos. Kim considera que hay un predominio de argumentos que sostienen el anomalismo psicofísico, esto es, la tesis de que no hay, ni puede haber, leyes que conecten estados intencionales con estados fisiológicos. Si esto es así, concluye, entonces la solución en términos de los equivalentes nómicos sería una posibilidad ociosa.

[13] Kim no menciona las razones que llevan a Goldman a rechazar que su posición pueda ser identificada con el epifenomenismo. Adviértase que, por otra parte, el ejemplo dado por Kim es algo problemático: el que se afirme que un síntoma *A** de una enfermedad *A* es un epifenómeno simultáneo depende de que se sostenga además de que la enfermedad se inició con (o simultáneamente a) la aparición de ese síntoma. Pero podría afirmarse que cualquier estado patológico se inicia *antes* de la aparición de cualquier síntoma, con lo que la posibilidad de epifenómenos simultáneos se esfuma. La dificultad radica aquí, como en otros problemas, en fijar un momento preciso de inicio de un determinado proceso o estado.

En tercer lugar, y suponiendo que existieran tales leyes de correlación, la situación en la cual tanto C* como C** son vistos como nomológicamente necesarios y suficiente el uno para el otro, y en la cual cada uno de ellos es un *explanans* completo e independiente para un mismo acontecimiento *E*, constituye una situación 'intrínsecamente inestable'. Esto es así, prosigue Kim, cuando C* y C** 'son cada uno de ellos miembros de un sistema de sucesos (o conceptos) tal que los dos sistemas a los cuales cada uno respectivamente pertenece muestran el tipo de conexiones nomológicas sistemáticas que Goldman imagina para lo psicológico y lo fisiológico' (1989a, p. 246). La inestabilidad que esta situación genera se traduce en una presión para encontrar una explicación aceptable de la relación entre C* y C**. La inestabilidad desaparece y se restablece el equilibrio cognitivo, concluye Kim, cuando la equivalencia entre las dos presuntas causas o explicaciones nomológicamente equivalentes es reemplazada por la identidad o por alguna relación de dependencia asimétrica.

Coincidimos con Kim en que el argumento de Goldman con el cual pretende evitar la objeción de que plantea una sobredeterminación global de la conducta humana resulta inconcluyente. El hecho de que el ejemplo de dos causas nomológicamente no conectadas 'difiera significativamente' del caso de los equivalentes nómicos no autoriza a pensar que esa es una razón suficiente para descartar el segundo como caso de sobredeterminación. La pregunta que Kim formula con respecto a los presuntos casos de sobredeterminación es pertinente para decidir la cuestión ('(1) ¿Habría acaecido E si *C* no hubiera acaecido? y (2) ¿habría acaecido E si *C** no hubiera acaecido? Si la respuesta a ambas preguntas es 'si', éste es un caso clásico de sobredeterminación' (1989a, p. 253)). Si la respuesta de Goldman fuese afirmativa en ambos casos, parece difícil evitar la consecuencia de que toda la conducta humana está sobredeterminada. Goldman podría evitarla afirmando que los equivalentes nómicos simultáneos, por definición, no pueden ser concebidos como actuando de manera independiente uno del otro, de forma tal que el interrogante antes mencionado no puede aplicárseles. Pero en este caso, ¿qué manera existe de saber si en realidad no estamos en presencia de una identidad? Si este fuese el caso, quedaría abierto el camino para una posterior reducción de la psicología mentalista a la neurofisiología. Si bien Goldman no está interesado en

argumentar en favor de la independencia de ambos tipos de explicación, no es seguro que aceptaría una posible reducción de la psicología en favor de la neurofisiología. La solución de los equivalentes nómicos simultáneos –pensamos que Kim coincidiría- debe tender a resolverse en un caso de sobredeterminación, un caso de dependencia asimétrica, o un caso de identidad. En el análisis de Kim juega entonces un rol fundamental la noción de que la coexistencia de explicaciones de un mismo *explananda* es una situación epistémicamente inadecuada, que debe resolverse de una u otra manera. Expondremos a continuación el análisis de Kim relativo a los casos posibles de causación de un suceso por otros.

De acuerdo con Kim, el principio de exclusión explicativa se plantea de manera más persuasiva con respecto a las explicaciones causales de sucesos individuales. Supóngase, dice, que la explicación *A* de un suceso *E* propone como causa del mismo a *A*, mientras que la explicación *B* del mismo suceso *E* propone como causa del mismo a *A**. En esta situación se pueden distinguir diversas posibilidades. Caso I: *A* es igual a *A**. En este caso, se observa que aquí hay una causa, no dos. En el caso de la causación psicofísica esta posibilidad toma la forma de la teoría de la identidad mente-cuerpo. Este caso, sostiene Kim, proporciona la manera más simple y posiblemente más satisfactoria para aliviar la tensión generada por la existencia de explicaciones rivales de un mismo *explanandum*. Caso II: *A* es distinta de *A** pero en algún sentido es 'reducible a' o 'sobreviniente' a partir de *A**. En este caso, sostiene Kim, no existen dos explicaciones causales *independientes* del mismo suceso. Ambas explicaciones podrían coexistir, ya que una de ellas depende de la otra, de manera reductiva o por superveniencia. Caso III: ni *A* ni *A** son, tomadas aisladamente, una 'causa suficiente' de *E*, aunque cada una es un componente necesario de una causa suficiente. En este caso no tenemos dos explicaciones completas, entendiendo por esto una explicación en la cual se especifica un conjunto suficiente de condiciones causales para la producción del *explanandum*. Caso IIIa: *A* es parte propia de *A**. En este caso *A*, en tanto explicación de *E*, no es en sí misma completa y tampoco independiente de *A**. Caso IV: *A* y *A** constituyen eslabones en una misma cadena causal que conduce de *A* a *A** y luego a *E*. En este caso una vez más no hay dos explicaciones causales independientes: el *explanans* de uno, *A*, depende causalmente del *explanans* de otro,

*A**. Caso V: *A* y *A** son, cada uno, una causa suficiente de *E*. Este es un caso de sobredeterminación causal. En estos casos, Kim considera que no es obvio como debería formularse una explicación de por qué o en qué forma se originó el acontecimiento que se debe explicar. No obstante, sugiere que es plausible pensar que 'el no mencionar alguna de las causas sobredeterminantes produce una imagen engañosa e incompleta de lo que aconteció, y que ambas causas deberían figurar en cualquier explicación *completa* del suceso' (1989a, p. 252. Cursivas del autor). Si esto es correcto, concluye, no es éste un caso en el cual sean posibles dos explicaciones completas e independientes del mismo acontecimiento.[14]

El análisis de estos casos, sostiene Kim, resulta útil para esclarecer el significado de 'completud' e 'independencia', referido a las explicaciones y crucial para el principio de exclusión explicativa. Kim considera que examinando los casos particulares que parecen posibilitar dos explicaciones diferentes de un mismo *explanandum* pueden encontrarse (y, más aún, existe una necesidad de buscar) razones para decir que las explicaciones no son independientes o al menos que una de ellas es incompleta. Dos explicaciones de un mismo suceso crean una tensión epistémica, que se disipa cuando se obtiene una explicación de cómo ellas (o las causas a las que ellas aluden) se relacionan entre sí.

Estas tesis de Kim serán relevantes, como veremos, a la hora de desarrollar una crítica sistemática al materialismo no reduccionista y de analizar los problemas que éste enfrenta al tener que presentar una visión plausible de la causación mental.

4. Tesis del materialismo no reduccionista

Tal como se ha comentado, no todas las formas de materialismo son vulnerables a las mismas objeciones relativas a su incapacidad para enfrentar el

[14] La posición de Kim con respecto a los casos de sobredeterminación causal y explicativa no resulta clara. En su (1987) sugiere que los casos de sobredeterminación explicativa, interpretados en un marco realista de la explicación científica, parecen llevar a contradecir el principio de exclusión. Esta afirmación implica que no cree que una explicación que aluda a una sola de las causas sobredeterminantes sea de alguna manera incompleta. No obstante, indica que el realista explicativo que desee salvar el principio de exclusión puede negar que existan genuinos casos de sobredeterminación. Pero en su tratamiento de esta cuestión, hasta donde llega nuestro conocimiento, Kim no se pronuncia ni en favor ni en contra de la existencia de tales casos (tampoco hay acuerdo en la literatura sobre el tema. Para posiciones opuestas *cfr*. Bunzl (1979) y Loeb (1974)). La cuestión acerca de si existen genuinos casos de sobredeterminación excede el marco de esta tesis. Por razones que luego veremos, no obstante, el análisis de los casos de sobredeterminación no jugará un rol fundamental en los problemas de la causación mental.

problema de la causación mental. La Teoría de la Identidad o el eliminativismo, por ejemplo, eluden ciertos problemas al tener que explicar la causación mental (aunque, por supuesto, enfrentan otra clase de dificultades). Las posiciones que sí se han revelado como directamente afectadas son aquellas variantes del materialismo que combinan esta ontología con una posición férreamente antirreduccionista. Se impone aquí describir con cierto detalle las tesis básicas de estas posiciones, para proceder luego a desarrollar los argumentos que muestran que son inadecuadas para dar respuesta a los interrogantes relativos a la causación mental. En el resto de este apartado nos basaremos en Kim (1993a), en el cual describe de manera precisa los rasgos fundamentales de las posiciones de esta clase.

De acuerdo con Kim, la visión del mundo que subyace a las doctrinas materialistas de la mente es la llamada 'visión estratificada del mundo'. Esta posición se caracteriza por considerar que el mundo está organizado como una estructura jerárquicamente estratificada de 'niveles' de entidades, donde cada uno de los cuales tiene propiedades características. Cada nivel de esta estructura tiene entonces dos componentes: un conjunto de entidades constitutivas y un conjunto de propiedades para ese dominio. Las entidades pertenecientes a un nivel dado están compuestas por entidades pertenecientes a niveles inferiores: esta es la relación mereológica de *ser parte de*. Admitido esto, resta responder como las *propiedades* características de un nivel se relacionan con las propiedades características de los niveles adyacentes.

Es en la respuesta a este interrogante donde surgen diversas concepciones opuestas sobre temas de metodología y metafísica, dice Kim, incluyendo especialmente al problema mente-cerebro. Mientras que la posición reduccionista sostiene que las propiedades características de entidades pertenecientes a cierto nivel son *reductibles a*, o *reductivamente explicables en términos de*, propiedades y relaciones características de niveles inferiores, las posiciones no reduccionistas y eliminativistas coinciden en afirmar que tales propiedades no son reductibles de la manera que los primeros sugieren. No obstante, no reduccionistas y eliminativistas discrepan acerca del *status* de estas propiedades: mientras que los primeros sostienen que estas son genuinas y reales propiedades de objetos y sucesos, los segundos opinan que son agregados inútiles que pueden ser eliminados de una correcta visión de la realidad.

Las posiciones materialistas no reduccionistas con respecto al problema mente-cuerpo podrían ser llamadas merecidamente, dice Kim, la 'concepción recibida' [*received view*]. Si bien tienen como atributo central el ser monistas fisicalistas (monismo de sustancias), reconocen la especificidad de lo mental, afirmando la existencia de *propiedades mentales* no reductibles a propiedades físicas. En este sentido, son compatibles con el llamado 'dualismo fisicalista de propiedades', es decir, un dualismo de atributos psicológicos y físicos. El admitir la distinción entre propiedades mentales y físicas ha permitido preservar una intuición irrenunciable para muchos autores –la idea de que existen propiedades mentales, las cuales poseen un carácter distintivo- evitando a la vez las insalvables dificultades que presenta el dualismo sustancialista. Dentro de una ontología materialista, observa Kim, se admite una ideología dualista.

La tesis ontológica básica del materialismo no reductivo es entonces que lo físico tiene una especie de predominio. Toda existencia concreta es física; no existen particulares no físicos, ni 'sustancias mentales' cartesianas. Esta tesis de primacía de lo físico ha sido defendida afirmando que, a pesar de su irreductibilidad, las propiedades mentales son, en algún sentido, 'dependientes' o 'determinadas' por propiedades físicas. La relación de dependencia o determinación se ha reflejado en dos ideas que han predominado dentro del programa materialista no reductivo: la idea de la 'superveniencia' y la de la 'realización física'.

Esta clase de materialismo acerca de lo mental descansa entonces, de acuerdo con Kim, sobre cuatro supuestos fundamentales:

1. [Monismo Físico] Todo particular concreto es físico.
2. [Antirreduccionismo] Las propiedades mentales no son reductibles a propiedades físicas.
3. [La Tesis de la Realización Física] Todas las propiedades mentales son físicamente realizadas; esto es, siempre que un organismo o sistema ejemplifica una propiedad mental *M*, tiene alguna propiedad física *P* tal que *P* realiza *M* en organismos de esa clase. (…)
4. [Realismo Mental] Las propiedades mentales son propiedades reales de objetos y sucesos; no son meramente ayudas útiles en la formulación de predicciones o maneras ficticias de hablar[15] (1993a, p. 344).

[15] Para Kim, este conjunto de supuestos configura una posición muy similar al emergentismo.

Con la admisión del supuesto de que sólo existen entidades físicas, esta forma de materialismo se compromete además con la tesis de la *clausura causal del mundo físico*: todo suceso físico requiere de una causa también física. Por otra parte, la tesis de la realidad de las propiedades mentales implica para Kim un postulado que, presume, sería aceptado por muchos materialistas no-reductivistas: *ser real es tener poderes causales* (*dictum* de Alexander). Esta característica de las propiedades mentales es el rasgo que permite distinguir a esta forma de materialismo del epifenomenismo. Para este último, las propiedades mentales pueden ser causadas, pero no pueden ser causas: no tienen ningún rol activo en la estructura causal del mundo.

Este conjunto de supuestos, que parece preservar lo mejor de ambos mundos (la prioridad ontológica de lo físico con la irreductibilidad y potencia causal de lo mental), da lugar a un argumento sencillo y poderoso, expuesto por Kim en este mismo artículo (1993a), que conduce al epifenomenismo de lo mental. En el apartado siguiente describiremos la versión más reciente de este argumento, desarrollada por Kim en su (1998).

5. El argumento de la superveniencia, o 'la venganza de Descartes'

Examinaremos ahora el principal argumento expuesto por Kim en contra del materialismo no reduccionista, al que denomina 'el argumento de la superveniencia, o la venganza de Descartes'. Este argumento, en opinión de Kim, nuestra que la superveniencia, por sí misma, conduce a serias dificultades con la causación mental. En este argumento juega un rol central, entonces, el supuesto de que la aceptación de la superveniencia define un fisicalismo mínimo; si esta suposición es correcta, aun las formas más débiles de fisicalismo tendrán que enfrentar el argumento de una u otra manera.

Tal argumento es construido por Kim con la forma de un dilema, el cual aparentemente lleva a la conclusión de que la causación mental es ininteligible. En esencia, sostiene, este argumento es el resultado de superponer la superveniencia mente-cuerpo con el problema de la exclusión causal.[16]

[16] No deja de resultar llamativo, y así lo han advertido algunos autores, que Kim, quien ha sido uno de los principales autores que han desarrollado el concepto de superveniencia como un intento de dar respuesta a los interrogantes planteados por el problema mente-cuerpo, considere que tal concepto no es parte de la solución, sino parte del propio problema.

Kim comienza estableciendo los dos cuernos del dilema:

(i) O la superveniencia mente-cuerpo tiene lugar o no lo tiene.

Recordemos que la tesis de la superveniencia mente cuerpo afirma que 'las propiedades mentales sobrevienen a partir de propiedades físicas en el sentido de que si algo ejemplifica cualquier propiedad mental *M* en *t*, entonces hay una propiedad física de base *P* tal que la cosa tiene *P* en *t*, y necesariamente cualquier cosa que posea *P* en *t* tiene *M* en ese momento' (p. 39). El segundo cuerno del dilema es formulado de la siguiente forma:

(ii) Si la superveniencia mente-cuerpo no tiene lugar, entonces no hay una manera visible de comprender la posibilidad de la causación mental.

Acerca de la razón de esta conexión entre superveniencia y causación mental, Kim afirma que la razón más simple y obvia que un fisicalista tiene para aceptar (ii) consiste en el compromiso con la *clausura causal del mundo físico*. Si la superveniencia mente-cuerpo no tiene lugar (esto es, si lo mental no está 'anclado' en el mundo físico), la causación de lo mental hacia lo físico infringe obviamente la clausura causal de lo físico. La superveniencia mente-cuerpo basa cada fenómeno mental en este último dominio, ya que provee un conjunto de condiciones físicas que son (al menos) nomológicamente suficientes para ellos y de cuya ocurrencia dependen. Un corolario de esto es la tesis, sostiene Kim, de que ningún fenómeno mental puede tener lugar, y ninguna propiedad mental puede ser ejemplificada, si no está presente una condición física de base apropiada. La superveniencia, por lo tanto, mantiene a los fenómenos mentales dentro del ámbito de lo físico; lo físico determina lo mental en el sentido de que éste no constituye un dominio ontológicamente independiente que introduce influencias causales desde fuera del dominio físico. Pero la situación es potencialmente más grave, en opinión de Kim: la superveniencia, lejos de constituir una parte de la solución al problema mente-cuerpo, podría ser en sí misma una parte del problema.

El argumento continúa de la siguiente forma:

(iii) Supóngase que un ejemplo de una propiedad mental *M* causa la ejemplificación de otra propiedad mental *M**.

Este es un caso de causación de lo mental a lo mental. Teniendo en cuenta las consideraciones anteriores sobre la superveniencia, se sigue

(iv) *M** tiene una base de superveniencia física *P**.

Surge aquí la siguiente pregunta: ¿de dónde proviene este ejemplo de *M**? O ¿cómo llega *M** a ser ejemplificada en esta ocasión? Hay, aparentemente, dos posibles respuestas a considerar:

(v) *M** es ejemplificada en esta oportunidad: a) porque, *ex hyphotesi*, *M* causa que *M** sea ejemplificada; b) porque *P**, la base de superveniencia física de *M**, es ejemplificada en esta oportunidad.

De acuerdo con Kim, debe percibirse aquí una tensión real entre las dos respuestas. Bajo el supuesto de la superveniencia mente-cuerpo, *M* tiene lugar porque su base de superveniencia *P** tiene lugar, y, en tanto *P** tenga lugar, *M** ocurrirá sin importar cuales otros sucesos precedan al ejemplo de *M**. Esto pone en riesgo la afirmación de que *M* es la causa de *M**; *P** por sí mismo parece completamente responsable de la ocurrencia de *M**. Debido a esto, el único modo en el cual algo puede tener un rol en la causación de *M** tendría que ser vía una relación con la base de superveniencia de *M**, esto es, *P**.

(vi) *M* causa *M** al causar *P**. Esta es la forma en la cual *M* causa que *M** sea ejemplificada en esta oportunidad.

En opinión de Kim, puede haber un principio plausible involucrado aquí, el cual es por sí mismo suficiente para justificar (vi): para causar la ejemplificación de una propiedad sobreviniente se debe causar la ejemplificación de la base (o una de las bases) de esa propiedad. Pero debe observarse aquí, continúa Kim, que (vi) afirma que una propiedad mental *M* causa que una propiedad física, *P**, sea

ejemplificada. Se trata de un caso de causación de lo mental a lo físico. Entonces lo que el argumento ha mostrado es lo siguiente: bajo el supuesto de la superveniencia mente-cuerpo, la causación de lo mental a lo mental implica, o presupone, la causación de lo mental a lo físico (causación descendente). La cuestión que se enfrenta aquí es, entonces, la de dar sentido a la causación de lo mental a lo físico, bajo la premisa de la superveniencia mente-cuerpo. Volviendo a (vi), vemos que bajo esta premisa se sigue que

(vii) *M* tiene su propia base de superveniencia *P*.

Dado lo anterior, continúa Kim, deben compararse *M* y *P* en relación con su *status* causal respecto de *P**. Ya sea en el caso en que la causación se considere basada en suficiencia nomológica, como en el caso en que se entienda esta relación en términos de contrafácticos, *P* parece calificar como causa de *P**. En el primer caso, dado que *P* es suficiente para *M* y *M* es suficiente para *P**, *P* es suficiente para *P**. En el segundo, si *P* no hubiera ocurrido *M* no habría ocurrido, y dado que si *M* no hubiese ocurrido *P** no habría ocurrido, puede razonablemente concluirse, observa, que si *P* no hubiera tenido lugar *P** tampoco habría ocurrido.

Parece observarse una superabundancia de causas: tanto *M* como *P* parecen elegibles como causas suficientes de *P**. No es posible, sostiene Kim, escapar a la amenaza de sobredeterminación sosteniendo que la situación involucra una cadena causal de *P* a *M* y luego a *P**, con *M* como un nexo causal intermediario; la relación entre propiedades de base y propiedades sobrevinientes no puede ser concebida adecuadamente como una relación causal. Por una parte, sostiene, la ejemplificación de las propiedades relacionadas es enteramente simultánea, mientras que se cree de manera standard que las causas preceden a sus efectos; en segundo lugar, es difícil, y quizás incoherente, imaginar un nexo causal con eslabones intermediarios entre las propiedades subvenientes y las sobrevinientes. Más aún, advierte, en el caso en estudio, la existencia de un nexo causal en el cual *M* es tomado como una causa no física de *P** violaría la clausura causal del mundo físico, opción que está vedada al fisicalista.

Tampoco es plausible, prosigue que *M* y *P*, juntas, constituyan una causa singular suficiente de *P**. Hay dos razones para esto. En primer lugar, *P* es por sí misma una causa suficiente para *P**, y también lo es *M*. Es difícil ver como *M* y *P* podrían aportar poderes causales adicionales que no posean *M* por sí misma y *P* por sí misma. En segundo lugar, este enfoque sólo es posible si se afirma que *M* es un componente necesario de la causación de *P**, y esto implicaría una violación del principio de clausura causal.

Por último, finaliza Kim, no es posible tomar este caso como uno de sobredeterminación causal, en el cual el ejemplo de *P** es causalmente sobredeterminada por dos causas suficientes, *P* y *M*. Esta posibilidad, además de tener la implausible consecuencia de hacer de todo caso de causación mental un caso de sobredeterminación, enfrenta dos clases de dificultades. En primer lugar, al admitir una causa física como posible sustituto para cada causa mental, convierte a éstas en prescindibles en cualquier caso. En segundo lugar, enfrenta nuevamente la amenaza de violación del principio de clausura causal del mundo físico.

Kim sostiene que la manera más natural de concebir la situación es la siguiente:

(viii) *P* causa *P**, y *M* sobreviene a partir de *P* y *M** a partir de *P**.

Esto explica, en su opinión, las regularidades observadas entre ejemplos de *M* y ejemplos de *M**. Estas regularidades no son accidentales; están claramente basadas en leyes y pueden ser capaces de apoyar contrafácticos apropiados. Sin embargo, advierte, si se comprende la diferencia entre procesos causales generativos y genuinos, por una parte, y regularidades no causales que son parasitarias de procesos causales reales, se estará en posición de entender el cuadro sugerido en (viii). En el caso de la supuesta causación *M*-*M**, la situación es similar a la de una serie de sombras proyectada por un automóvil en movimiento: no hay conexión causal entre la sombra del auto en un instante y su sombra en un instante posterior; cada una de ellas es simplemente un efecto del movimiento del auto. Por lo tanto,

(ix) Las relaciones causales de *M* a *M** y de *M* a *P** son sólo aparentes; surgen de un proceso causal genuino de *P* a *P**.

De dónde el dilema

(x) Si la superveniencia mente-cuerpo no tiene lugar, la causación mental es ininteligible; si tiene lugar, la causación mental es nuevamente ininteligible. Por lo tanto, la causación mental es ininteligible.

Este argumento, concluye Kim, plantea serias dificultades a toda forma de materialismo no reduccionista que se encuentre comprometido con la existencia de propiedades físicas y psicológicas. En nuestra opinión, este argumento es sumamente fuerte y de serias consecuencias para quienes deseen sostener formas de materialismo con las características indicadas.

Antes de exponer los restantes argumentos que, con variantes, desembocan en la ineficacia causal de las propiedades mentales, nos parece de interés analizar una comentada doctrina expuesta por John Searle en su libro *El redescubrimiento de la mente* (1992): el naturalismo biológico. Esta doctrina es analizada y criticada por Kim en su (1998), como ejemplo de una de las diversas posiciones que se ven seriamente cuestionadas por el argumento de la superveniencia. Sin embargo, nuestro interés al comentar la posición de Searle es principalmente el de poner de manifiesto la agudeza del argumento, cuestionando los intentos 'fáciles' de solución al problema.

6. Searle y el 'naturalismo biológico'

De acuerdo con el naturalismo biológico, sostiene Searle, los fenómenos mentales, con todos sus rasgos subjetivos y cualitativos, no se diferencian en principio de otros fenómenos biológicos, como la digestión o la reproducción. En cuanto a la relación entre los fenómenos mentales y los biológicos, Searle sostiene que los primeros son *causados* por los segundos, supuestamente por procesos neurales en el cerebro. A la vez, tales fenómenos mentales *son rasgos del cerebro*. Searle adopta una posición netamente realista con respecto a lo mental, y defiende su eficacia causal; en especial, pone énfasis en la eficacia causal de la conciencia. Esta

solución es, en su opinión, muy simple, y ha estado al alcance de cualquier estudioso aproximadamente desde hace un siglo.

En su (1995) Kim expone una crítica a nuestro entender contundente a la propuesta de Searle, crítica que repetirá, con algunos matices, en su (1998). En particular, sostiene que esta propuesta es vulnerable al argumento de la superveniencia. El análisis de la posición de Searle resultará ilustrativo de las dificultades que plantean las soluciones 'fáciles' al problema, y nos permitirá exponer de manera clara algunas de sus debilidades. Describiremos sucintamente a continuación las objeciones planteadas por Kim en los trabajos mencionados, para luego profundizar en un aspecto de la propuesta de Searle que no está presente en tal análisis.

En opinión de Searle, la causa de que la mayoría de los filósofos no haya reparado en la sencilla solución que propone para el problema reside en que tales filósofos se han mantenido dentro del marco de la 'metafísica cartesiana'. Tanto los fisicalistas como los funcionalistas han quedado atados, según Searle, en esta visión anticientífica y anticuada. En su reemplazo, Searle propone una visión de la realidad consistente en un esquema de múltiples estratos o 'niveles' de objetos y propiedades, modelo que, según Kim, ha devenido en una concepción familiar en la literatura sobre la reducción teórica. Con respecto a la cuestión crucial relativa a la manera en que los objetos y las propiedades de los distintos niveles se relacionan entre sí, y en especial en lo que se refiere a la relación entre estados mentales y estados del cerebro, su principal afirmación es que los procesos neuronales *causan* los fenómenos mentales. Searle denomina a esta clase de relación 'superveniencia causal'. Podemos suponer, comenta Kim, que ésta es la relación que Searle considera como válida acerca de las relaciones entre propiedades de distintos niveles: las propiedades de nivel superior son sobrevinientes causalmente a partir de propiedades del nivel inferior. Esta concepción, prosigue Kim, conduce naturalmente a Searle a afirmar que las propiedades y fenómenos de nivel superior son *explicables* por propiedades del nivel inferior. Pero si bien Searle hace uso de expresiones como 'superveniencia' o 'emergencia', comentará Kim luego, este uso es sumamente idiosincrásico. Donde otros autores sostendrían que los fenómenos

mentales sobrevienen, o emergen, de sus sustratos neurales, Searle afirmaría que tales fenómenos mentales son *causados* por los procesos de nivel inferior.

Si las afirmaciones de Searle relativas a que lo mental es físico, prosigue Kim, significaran solamente que las propiedades mentales son rasgos de nivel superior del cerebro, resultaría dudoso que alguien estuviese en desacuerdo. Sin embargo, ésta es meramente una maniobra verbal que deja sin respuesta a un interrogante metafísico fundamental: ¿son esos rasgos de nivel superior del cerebro reductibles a, o reductivamente identificables con, las propiedades de nivel inferior a partir de las cuales sobrevienen? La respuesta de Searle, advierte Kim, es una negativa rotunda. Pero esto es lo que define al dualismo de propiedades, doctrina con la que Searle no desea identificarse.

¿Puede sostener Searle, dado el conjunto de sus compromisos, los amplios poderes causales que atribuye a los estados mentales? La respuesta de Kim es negativa. Presenta para esto una variación del argumento de la superveniencia ya expuesto.

Considérese, propone Kim, un suceso mental, una ejemplificación de alguna propiedad mental *M*. Este suceso, bajo el enfoque de Searle, es causado por la ejemplificación de alguna propiedad biológica B. Supongamos que *M* tiene poderes causales, poderes para causar la ejemplificación de otras propiedades. Podemos distinguir dos casos: (i) la propiedad cuya ejemplificación puede ser causada es en sí misma una propiedad mental, (ii) la propiedad es una propiedad física. (ii) es por supuesto un caso de causación de lo mental a lo físico ('causación descendente'); (i) puede ser denominado 'causación en el mismo nivel'. Searle, puede presumirse, desea mantener ambas. Consideremos primero la posibilidad (i): un ejemplo de *M* causa la ejemplificación de otra propiedad mental *M**. Debemos recordar, sin embargo, que bajo el enfoque de Searle el ejemplo de *M**, así como el ejemplo de *M*, son también causados por algún fenómeno biológico subyacente, un ejemplo de alguna propiedad biológica *B**. Pareciera que el ejemplo de *M** tendría *dos causas suficientes distintas*, un fenómeno mental (*M*), y un fenómeno biológico (*B**), y estaría, por lo tanto, causalmente sobredeterminado. Esto, desde luego, puede generalizarse rápidamente: todo caso de causación de lo mental a lo mental involucra la sobredeterminación del efecto. Y dado el hecho de que cada suceso mental tiene

una causa suficiente en un proceso biológico, se puede preguntar acerca del significado, o necesidad, de la causa mental.

Searle, sostiene Kim, está muy cerca de reconocer este problema general. Pero si bien admite la existencia de distintos tipos de explicación de un mismo fenómeno (de lo macro a lo macro, de lo micro a lo micro y de lo micro a lo macro) no explica cómo es posible la presencia de estas tres relaciones causales al mismo tiempo.

Hemos visto que al aplicar el argumento de la superveniencia al naturalismo biológico de Searle queda abierta la posibilidad de que la ejemplificación de *M** sea causalmente sobredeterminada: la ejemplificación de *M** es causada por la ejemplificación de *M*, o es causada por la ejemplificación de la propiedad neuronal *P**. A objeciones de esta clase, observa Kim (1998), Searle ha respondido de la siguiente forma:

> ¿Implica esto una sobredeterminación? En absoluto. El mismo sistema es descripto en diferentes niveles (...) En breve, el mismo sistema admite diferentes descripciones causales en diferentes niveles, todos los cuales son consistentes, y ninguno de ellos implica sobredeterminación o violación de la clausura causal.
> Ahora tengo, supongamos, una sensación conciente de dolor. Esta es causada por patrones de activación neuronal y es realizada en el sistema de neuronas. Supongamos que el dolor causa el deseo de tomar una aspirina. El deseo es también causado por patrones de activación neuronal y es realizado en el sistema de neuronas (...) Puedo decir correctamente que tanto mi dolor causa mi deseo cuanto que las secuencias de activación neuronal causan otras secuencias. Esas son dos descripciones diferentes pero consistentes del mismo sistema dadas a diferentes niveles (Searle, 1995, p. 219, citado en Kim, 1998, pp. 48-49).[17]

Decir que 'el dolor causa el deseo de tomar una aspirina' y que 'la activación neural *P* causa la activación neural *P* son descripciones de la misma situación*, señala Kim, es plausible sólo si se está dispuesto a admitir que 'dolor' y 'activación neural *P*' son descripciones del mismo fenómeno (en diferentes niveles). Lo mismo ocurre con 'deseo de aspirina' y 'activación neural *P**'.

Kim advierte que no considera que ésta sea una alternativa implausible; más aún, alguna forma de reduccionismo de esta clase puede muy bien resultar la única manera viable en la cual los fenómenos mentales tengan un rol causal genuino en el

mundo físico. Sin embargo, en relación con la propuesta de Searle, las identidades psiconeurales de esta clase ponen en riesgo sus afirmaciones de que la activación neural *P causa* el dolor, y de que la activación neural *P* causa* el deseo de aspirina. Si estas afirmaciones causales se sostienen, entonces 'dolor' y 'activación neural *P*' no pueden ser descripciones del mismo fenómeno, y lo mismo ocurre con el otro par de afirmaciones. Si bien Searle afirma que esas son descripciones 'en distintos niveles', ¿qué podría significar esto? Searle necesita, concluye Kim, formular una ontología y un lenguaje de la causación razonables para tornar sus afirmaciones acerca de la relación mente-cuerpo inteligibles y consistentes.

Las críticas expuestas por Kim constituyen, a nuestro entender, una refutación concluyente de la pretensión de Searle de haber resuelto de manera sencilla el problema mente-cuerpo. En alguna medida, la posición de Searle constituye un buen ejemplo de un filósofo que no toma en serio el problema de la causalidad mental; su 'solución' al problema no representa ningún avance real en la comprensión de la relación entre lo mental y lo físico.

Sin embargo, consideramos que las dificultades de la propuesta de Searle no se agotan en lo expuesto por Kim. No discutiremos aquí el punto de vista de Searle según el cual las relaciones entre un conjunto de entidades (por caso, átomos de hidrógeno y oxígeno) y otro conjunto de entidades (las moléculas de agua), son relaciones de tipo causal, aunque, puede hacerse notar, la divergencia con el punto de vista mayoritario de que la relación causal es una relación entre sucesos y no entre conjuntos de entidades o propiedades merecería bastante más reflexión que la que el autor le dedica. Tampoco analizaremos sus afirmaciones relativas a la emergencia y a la superveniencia. Sí ampliaremos brevemente el análisis de su respuesta al problema de la competencia causal entre sucesos mentales y físicos.

Consideramos que Kim está en lo correcto al advertir que la afirmación de identidad pone en riesgo a la afirmación de causalidad: si *P* es la causa de *M*, no parece plausible afirmar que a la vez *P* es idéntica a *M*. Sin embargo, a nuestro entender Kim descuida el hecho de que, para Searle, las relaciones entre identidad y

[17] Searle, John (1995), 'Consciousness, the Brain and the Connection Principle: a Reply', *Philosophy and Phenomenological Research*, 55.

causalidad son tan idiosincrásicas como su uso de la noción de causalidad. Searle contesta lo siguiente con respecto a las objeciones a su posición relativa a las relaciones entre identidad y causalidad:

> A veces, mis puntos de vista encuentran resistencia a causa de una concepción equivocada de las relaciones entre causalidad e identidad. U. T. Place (1988), por ejemplo, escribe: 'De acuerdo con Searle, los estados mentales son, a la vez, idénticos a, y causalmente dependientes de, los correspondientes estados cerebrales. Mi posición es que no es posible nadar y guardar la ropa. O bien los estados mentales son idénticos a los estados cerebrales o unos dependen causalmente de los otros. No pueden ser ambas cosas' (p. 209).
>
> Place piensa en casos como 'Estas huellas pueden depender causalmente de los zapatos del ladrón, pero no pueden ser, a la vez, idénticas a estos zapatos'. Pero ¿qué pasa con 'El estado líquido de esta agua puede ser causalmente dependiente de la conducta de las moléculas, y también puede ser un rasgo del sistema que está compuesto por las moléculas'? Me parece igualmente obvio que mi presente estado de conciencia está causado por la conducta neuronal de mi cerebro y que ese mismo estado es sólo un rasgo de nivel superior del cerebro. Si esto quiere decir nadar y guardar la ropa, nademos (p. 102, *n.* 4).

La pobreza de esta réplica resulta asombrosa. Aun admitiendo el punto de vista de Searle –por otra parte, minoritario- según el cual la relación causal no sólo tiene lugar entre sucesos, sino también entre conjuntos de entidades y propiedades, nos hallamos siempre con una relación entre dos cosas (sucesos, entidades o propiedades). Pero en el caso de la identidad –y si en algo se ha enfatizado en la literatura filosófica sobre el problema mente-cuerpo en las últimas décadas es en que la teoría de la identidad de tipos pretendía resolver el problema al postular no dos clases de fenómenos o sustancias, sino sólo una- se trata de una sola cosa, y no de dos. Según Searle, al parecer, la causa podría en ciertos casos ser idéntica al efecto. Searle no se toma el trabajo de explicar cual es la 'concepción equivocada de las relaciones entre causalidad e identidad'; simplemente se limita a afirmar que es 'obvio' que un proceso mental puede ser causado por un proceso neuronal y ser a la vez idéntico a éste. En nuestra opinión, su solución al problema de la competencia causal entre sucesos mentales y físicos agrava las dificultades de su propuesta en vez de disminuirlas.

Por supuesto, las relaciones metafísicas tales como la causalidad y la dependencia mereológica son nociones sumamente controvertidas, sobre las que se

está lejos de alcanzar un consenso. Por ejemplo, algunos filósofos desconfían de la superveniencia, bajo la sospecha de que no sea distinta de una relación de identidad.[18] No obstante, esta es una cuestión muy compleja, como todas las discusiones que, sobre la superveniencia, se han desarrollado en los últimos años. Pero pretender que se acepte sin mayor argumentación, como Searle lo hace, que pueden existir casos de relaciones entre sucesos en los cuales uno es causado por el otro y a la vez son idénticos, es inadmisible.

Searle presenta a su posición como algo superior a las propuestas tradicionales, a la vez que afirma que es distinta de ellas.[19] Recordemos aquí la afirmación central de Searle con respecto al problema mente-cuerpo:

> El famoso problema mente-cuerpo (...) tiene una solución muy simple. Esta solución ha estado al alcance de cualquier persona culta desde que empezaron a realizarse, hace más o menos un siglo, trabajos serios sobre el cerebro y, en un sentido, todos sabemos que es verdadera. Tal solución es la siguiente: *los fenómenos mentales están causados por procesos neuropsicológicos del cerebro y son a su vez rasgos del cerebro*. Para distinguir este punto de vista de muchos otros que existen en el mercado lo llamaré 'naturalismo biológico' (1992, p. 15. Cursivas nuestras).

Con esta posición[20] Searle pretende haber logrado una concepción superadora de lo que considera son serios errores de los puntos de vista alternativos, llámense estos materialismo, monismo, dualismo de sustancias o dualismo de propiedades. En vista de las dificultades que enfrenta su posición, y la carencia de argumentos plausibles para defenderla, no puede más que concluirse que ha estado muy lejos de lograr su objetivo.

[18] Por ejemplo, Ruben (1990).

[19] 'Mi punto de vista, quiero subrayarlo, no es una forma de dualismo. Rechazo tanto el dualismo de propiedades como el dualismo de sustancias; pero precisamente por las mismas razones por las que rechazo el dualismo, rechazo también el materialismo y el monismo. El gran error es suponer que se debe elegir entre esos dos puntos de vista' (p. 42).

[20] Reiterada a lo largo del libro en relación con fenómenos mentales particulares: '*la conciencia es un rasgo biológico de los cerebros humanos y de ciertos animales. Está causada por procesos neurobiológicos y es una parte del orden biológico natural como cualquier otro rasgo biológico, como lo son la fotosíntesis, la digestión o la mitosis*' (p. 102. Cursivas del autor).

7. Relevancia causal y eficacia causal: las propiedades de nivel superior en peligro

Un análisis diferente al de Kim, pero cuyas consecuencias son muy similares para la posibilidad de la causación mental, es expuesto por Jackson y Pettit (1990). Estos autores tratan de probar que, dadas ciertas condiciones, las propiedades de nivel superior (entre las que se encontrarían las propiedades mentales), serían causalmente ineficaces. Para estos autores, sin embargo, hay una manera en la cual las propiedades pueden ser *causalmente relevantes*, y por lo tanto participar en explicaciones causales, sin ser causalmente eficaces.

Para Jackson y Pettit, el problema se plantea al admitir tres suposiciones plausibles acerca de las explicaciones causales, las cuales, al ser combinadas con un cuarto supuesto que parece igualmente aceptable, generan un serio problema para el rol de las propiedades de las ciencias especiales en tales explicaciones. Estas tres suposiciones primarias son las siguientes:

1. Una explicación causal de algo debe orientarnos hacia las propiedades causalmente relevantes del factor identificado como explicativo;
2. Un modo en que las propiedades son causalmente relevantes es siendo causalmente eficaces. Una propiedad es causalmente eficaz si es una propiedad en virtud de cuya ejemplificación, al menos en parte, el efecto ocurre;
3. Una propiedad F no es causalmente eficaz en la producción de un efecto *e* si las tres condiciones siguientes están conjuntamente satisfechas:
   a. Hay una propiedad distinta G tal que F es eficaz en la producción de *e* si y sólo si G es también eficaz en su producción;
   b. El ejemplo de F [*F-instance*] no ayuda a producir el ejemplo de G [*G-instance*], si G es eficaz, ayuda a producir *e*; F y G no son factores causales secuenciales;
   c. El ejemplo de F no se combina con el ejemplo de G, directamente o por medio de efectos posteriores, para ayudar a producir *e* (ni viceversa): F y G no son factores causales coordinados.

Un ejemplo en el que se satisfacen simultáneamente las condiciones 3.a, 3.b y 3.c es el siguiente: un cristal se golpea y se rompe. Una primera explicación posible es que se ha roto a causa de su fragilidad. Una segunda explicación refiere a la

estructura molecular particular del cristal. En este caso, la propiedad de la fragilidad es eficaz en producir la rotura si y sólo si la propiedad de una determinada estructura molecular está presente (satisface 3.a). La propiedad de la fragilidad no ayuda a producir la propiedad de una determinada estructura molecular; no hay un lapso de tiempo entre la producción de ambas propiedades (satisface 3.b). Por último, la fragilidad no se combina con la estructura para la producción de *e* (satisface 3.c).

En casos como el descripto, parece evidente que F no puede ser considerada eficaz en el sentido en que lo es G. La relación entre la ejemplificación de F y la ocurrencia de *e* es secundaria a la relación entre la ejemplificación de G y la ocurrencia de *e*.

A las tres suposiciones descriptas puede adicionársele el principio siguiente:

4. La única manera en que una propiedad puede ser causalmente relevante para la producción de un efecto es siendo causalmente eficaz en el proceso de su producción.

Con este agregado, las consecuencias son altamente negativas para el rol explicativo de las propiedades que no se encuentren dentro del dominio de una ciencia básica (presumiblemente, la física), ya que sólo en este dominio se podría hallar propiedades que escaparían a la irrelevancia causal derivada de la combinación de los supuestos 1-4. La conjunción de estos cuatro supuestos llevaría, según Jackson y Pettit, a desechar las explicaciones de las llamadas 'ciencias especiales' y también las del sentido común, que serían reemplazadas por explicaciones dentro de los dominios de una tal disciplina. Propiedades tales como la cohesión de un grupo, la propiedad de un estado mental de ser una creencia de que *p*, o la propiedad de ser un rasgo biológico que maximiza el ajuste, no podrían ser invocadas en explicaciones causales.

Si bien el desechar las propiedades invocadas en las explicaciones de sentido común y en las ciencias especiales podría agradar a algunos, advierten, no sólo tales explicaciones corren el riesgo de ser abandonadas. Tampoco las explicaciones que hagan uso de la cuantificación existencial (que implican la referencia a un indeterminado) serían explicaciones apropiadas, limitación que abarcaría explicaciones no sólo de las ciencias especiales, sino también explicaciones físicas. Por ejemplo, supóngase que se explica el ruido hecho por algún mecanismo

recurriendo a la propiedad del mecanismo de que alguna de sus partes está floja. Esa propiedad se relaciona con otra propiedad más específica de la misma manera en que F se relaciona con G: una parte determinada del mecanismo está floja. Si esto fuese así, tampoco podría ser una explicación adecuada, por ejemplo, afirmar que la radiación emitida por un trozo de uranio se debe al hecho de que algunos de sus átomos están decayendo.

La renuncia al conjunto de explicaciones que quedaría descartado de aceptarse los cuatro supuestos descriptos, en opinión de Jackson y Pettit, es una consecuencia indeseable. En particular, tales explicaciones nos brindan información que no está disponible sólo por el hecho de tener acceso a las explicaciones correspondientes que hagan referencia a propiedades de nivel inferior. Consideran que existe una forma de salvar su relevancia causal, ya que no su eficacia causal, y tal forma se logra por medio del rechazo al cuarto supuesto, que parece el más cuestionable. Pospondremos el examen de esta propuesta hasta el capítulo IV.

8. ¿Conduce el funcionalismo al epifenomenismo?

Un tercer planteo reciente en el cual se pone en tela de juicio la eficacia causal de ciertos estados mentales es debido a Block (1990). En este artículo Block plantea el problema de si el funcionalismo puede evitar el epifenomenismo, entendiendo esta doctrina como la posición que sostiene que 'lo que pensamos o queremos no tiene relevancia causal para lo que hacemos' (p. 29).[21] Block formula inicialmente el problema en términos de lo que denomina 'paradoja de la eficacia causal del contenido', la cual surge a partir de la aceptación de las siguientes afirmaciones, que parecen verdaderas pero que resultan incompatibles: 1) el contenido intencional de un estado intencional es causalmente relevante para la conducta; 2) el contenido intencional de esos estados se reduce a los significados de representaciones internas, y 3) los procesadores internos son sensibles a las 'formas sintácticas' de las representaciones internas y no a sus significados. Sin embargo, este conjunto de afirmaciones, cuya consecuencia parece ser el epifenomenismo de lo mental, tiene un alcance limitado a sistemas simples, no para sistemas genuinamente

intencionales como los seres humanos, ya que la afirmación 3) no es aplicable a ellos. En tales sistemas genuinamente intencionales las diferencias en el significado de las representaciones conducen a diferencias conductuales, lo cual no ocurre con sistemas más simples.

Pese a que el funcionalismo puede enfrentar satisfactoriamente este primer problema, no ocurre lo mismo con un segundo problema que parece conducir nuevamente al epifenomenismo.

Este segundo problema es planteado por Block en los siguientes términos: las propiedades funcionales (esto es, la clase de propiedades que son esenciales para el enfoque funcionalista) son causalmente inertes en ciertos casos cruciales. Las propiedades funcionales son propiedades que consisten en la posesión de una u otra propiedad (esto es, ciertas propiedades no funcionales), que tienen ciertas relaciones causales con otras y con inputs y outputs. En la producción de esos outputs, son las propiedades no funcionales las que son causalmente relevantes, no las propiedades funcionales. Block proporciona algunos ejemplos, de los cuales transcribiremos uno. Considérese, sugiere, a la 'dormitividad', construida como una propiedad de segundo orden, esto es, la posesión de cierta propiedad u otra (por ejemplo, una propiedad química de primer orden) que es causalmente relevante para el dormir. Esto es, *x* es dormitivo = *x* tiene cierta propiedad que es causalmente relevante para el dormir. Si una píldora para dormir fuese introducida en la comida de una persona sin que ésta lo supiera, la propiedad de la píldora que es causalmente relevante para que la persona se durmiera sería la propiedad química (presumiblemente de primer orden) y no, parecería, la dormitividad por sí misma. Diferentes sustancias dormitivas actuarán por medio de diferentes propiedades químicas, una en el caso de que se trate de Valium, otra en el caso de que se trate de Seconal. Pero a menos que la persona sepa acerca de la dormitividad de la píldora, pregunta Block, ¿cómo podría la dormitividad por sí misma ser causalmente relevante para el hecho de que la persona se durmiera?

Por supuesto, señala, si la persona *sabe* acerca de la dormitividad, entonces puede ser causalmente relevante para el dormir; es el bien conocido fenómeno del

[21] Utilizando el término 'epifenomenismo', como él mismo advierte, en un sentido distinto del usual, el cual hace referencia al hecho de que los estados mentales pueden ser causados, pero no pueden ser

efecto placebo. Más aún, podría haber píldoras dormitivas que funcionen sin ningún efecto de primer orden, píldoras que requieran el reconocimiento de su carácter para lograr la dormitividad.

No obstante, permanece en pie el hecho de que las propiedades de segundo orden no son *siempre* causalmente relevantes para los efectos en términos de los cuales son definidas. Los únicos casos en los cuales las propiedades de segundo orden parecen ser causalmente eficaces son aquellos en los cuales un ser inteligente las reconoce como tales. Pero no son estos los casos en los que Block piensa cuando habla acerca de 'casos standard'; por el contrario, estos casos son aquellos en los cuales la propiedad de segundo orden es definida en términos de un efecto, y tal efecto es producido sin reconocimiento alguno de la propiedad de segundo orden por un ser inteligente.

Así como al inicio del artículo Block plantea la 'paradoja de la eficacia causal del contenido' como un conjunto de afirmaciones inconsistente, el segundo argumento en contra del funcionalismo puede ser sintetizado en otro conjunto de afirmaciones inconsistente: 1) las propiedades de las ciencias especiales son causalmente relevantes para los efectos que esas ciencias explican y predicen; 2) las propiedades de tales ciencias especiales son a menudo funcionales; 3) las propiedades funcionales son, de manera standard, causalmente irrelevantes para los efectos en términos de los cuales son definidas.

El conjunto de argumentos que hemos expuesto en el presente capítulo parece conducir, sin excepciones, a dudar seriamente acerca del rol causal de las propiedades mentales (y, en general, de cualquier propiedad que no pertenezca al dominio de la física básica). No hay duda de que tales argumentos son sólidos –muchas veces más sólidos que los que se ofrecen como réplica-; sin embargo, y dadas las consecuencias que el abandono de la causación mental implicaría, no deberá resultar extraño encontrar una batería de posiciones que, de una u otra manera, han intentado lidiar ya sea con los argumentos expuestos, señalando insuficiencias reales o presuntas en ellos, o, más modestamente, con algunas de sus

causas; son causalmente inertes.

indeseables consecuencias. A varias de estas cuestiones nos referiremos en los capítulos que siguen.

CAPÍTULO II: CONSECUENCIAS Y ESTRATEGIAS ANTE EL PROBLEMA

1. La exclusión: qué implica y cómo enfrentarla

En el capítulo anterior hemos descripto algunos de los principales argumentos que amenazan la eficacia causal de las propiedades mentales. Tales argumentos suscitan algunas preguntas básicas. Si se considera que son correctos y el materialismo no reduccionista enfrenta una inconsistencia interna insubsanable, deben determinarse sus consecuencias para toda teoría de lo mental (y también para el sentido común) que presuponga que el dominio de lo mental depende del dominio físico pero posee poderes causales autónomos. Si se considera que los argumentos son erróneos, debe determinarse dónde se encuentra la premisa (o premisas) erróneas, y la forma en que puede neutralizarse su potencial destructivo para el materialismo no reduccionista.

Por lo tanto, en el presente capítulo abordaremos dos cuestiones principales. En primer lugar, las consecuencias del problema; los argumentos expuestos, conjuntamente con tesis que parecen plausibles, parecen tener consecuencias adicionales funestas para el programa del materialismo no reduccionista. En segundo lugar, describiremos las estrategias empleadas para lidiar con él. En el curso de este análisis, examinaremos someramente la conexión del problema con la investigación empírica en dos sentidos distintos: primero analizaremos la posibilidad de que la evidencia experimental pueda constituir un criterio para la elección de una u otra doctrina acerca de lo mental; luego intentaremos establecer las consecuencias que los argumentos de la exclusión tienen para enfoques metateóricos que han resultado fructíferos para el desarrollo de la investigación fáctica, en particular, el concepto de niveles explicativos ampliamente utilizado en psicología cognitiva e inteligencia artificial. Para la descripción de las consecuencias del problema y las estrategias empleadas para enfrentarlo seguiremos la descripción que proporciona Sabatés (2001).

2. Consecuencias del problema

Sabatés distingue básicamente tres tipos de consecuencias del argumento de la exclusión: el irrealismo de lo mental, la irrelevancia explicativa de lo mental y la ineficacia de las propiedades funcionales.[22] Examinaremos las dos primeras, poniendo especial énfasis en el problema de la irrelevancia explicativa. Los resultados del análisis de este último problema serán importantes al momento de evaluar el lugar que las explicaciones psicológicas pueden jugar dentro de un marco epifenomenista.

2.1. El irrealismo de lo mental

El irrealismo mental es la doctrina según la cual no hay entidades (objetos, propiedades, o cualquier otra cosa) que puedan ser categorizadas como mentales. Un criterio mayoritario (si bien, observa Sabatés, no unánime) para determinar cuando una propiedad o clase de propiedades es real es el criterio del poder causal:

> (CPC) 'Una propiedad es real sólo si contribuye a los poderes causales activos del objeto que la posee'(p. 26).

Dado este principio parecería que una consecuencia inevitable del argumento de la exclusión es el irrealismo de lo mental. Esta consecuencia, observa Sabatés, es más dañina aun para el materialismo no reduccionista que la ineficacia causal, ya que contradice una de sus afirmaciones básicas. Sin embargo, considera que el irrealismo no es una consecuencia necesaria del argumento de la exclusión; su aceptabilidad puede ser cuestionada a partir de la oposición al criterio causal para la realidad de las propiedades. Por una parte, observa, los argumentos en favor de este principio, principalmente epistemológicos, no son concluyentes; en segundo lugar, los

[22] Hemos hecho referencia a la dificultad suscitada para las propiedades funcionales al describir en el capítulo anterior el argumento de Jackson y Pettit (1990). Para Sabatés el problema de la ineficacia de las propiedades funcionales tal como es planteado por Jackson y Pettit es una versión especial del problema de la exclusión, con la diferencia de que estos autores sólo analizan el caso de la causación de lo mental a lo físico (aunque podría mostrarse a través de los mismos supuestos la ineficacia completa de lo mental). Una segunda diferencia (que resulta ser, en opinión de Sabatés, sólo aparente) consiste en que el problema planteado por Jackson y Pettit parece afectar a toda propiedad funcionalmente caracterizada, mientras que el problema de la exclusión parece afectar únicamente a las propiedades mentales. Pero esto es así sólo en apariencia, ya que el argumento de la exclusión puede generalizarse de modo tal de abarcar a toda propiedad sobreviniente (ya sean éstas biológicas o químicas), por lo que la diferencia de alcance entre ambos problemas no es tal.

argumentos más fuertes de esta clase no descartan la realidad de efectos causales que carecen de eficacia por sí mismos. Un criterio causal más amplio podría afirmar:

> (CRC) 'Una propiedad es real si contribuye a los poderes causales *activos o pasivos* del objeto que la posee' (p. 26. Cursivas del autor).[23]

Este criterio permitiría considerar a las propiedades epifenoménicas como propiedades reales, y sería suficiente, observa, para tornar compatibles al realismo mental y a la ineficacia causal.[24] En conclusión, sería un error suponer sin mayor análisis que la exclusión implica necesariamente el irrealismo acerca de lo mental. Una conclusión similar se observará con respecto a la irrelevancia explicativa.

2.2. Irrelevancia explicativa

En opinión de Sabatés, una segunda consecuencia dañina del problema de la exclusión es la potencial irrelevancia de las explicaciones psicológicas. Esta segunda consecuencia se origina por dos vías distintas: a) si al resultado de la ineficacia causal de lo mental se le agrega la tesis de que toda explicación es causal, resulta inevitable concluir que lo mental no puede ser explicativo. Esta consecuencia se basa en el supuesto de que toda explicación debe 'rastrear' una relación objetiva entre los sucesos que son descriptos por el *explanans* y el *explanandum*, y que esta relación objetiva es la relación causal; b) si, como ha resultado aceptable desde Davidson (1963), las explicaciones racionalizadoras son explicaciones causales, se llega por

[23] Sin embargo, Sabatés observa que este principio quizás no sea completamente adecuado: 'Aún (CRC) es tal vez demasiado fuerte. Supóngase que una propiedad B está causalmente aislada pero depende o sobreviene no causalmente a partir de una propiedad causalmente eficaz A. ¿No deberíamos considerarla parte de una red de relaciones reales y por lo tanto una propiedad genuina mientras evitamos enfoques más pragmáticos acerca de qué considerar una propiedad genuina?' (p. 39 *n*. 32).

[24] Armstrong (1978) menciona esta posibilidad, formulando a la vez una advertencia: 'Hay que notar, de paso, que parece posible concebir propiedades que proporcionen a los particulares sólo poderes activos o sólo poderes pasivos. Propiedades del primer tipo serían, por así decir, "motores inmóviles", en tanto que las propiedades del segundo tipo serían "epifenoménicas". Un problema de suma dificultad sería el de *si podríamos tener jamás razón para creer en la existencia de tales calles causales de un solo sentido en la esfera de las propiedades*. Pero, afortunadamente, nuestros propósitos actuales no parecen requerir que lleguemos a una decisión acerca de la cuestión' (p. 230. Las cursivas son nuestras). Parece plausible conjeturar que Armstrong está pensando precisamente en las propiedades mentales cuando se refiere a las propiedades 'epifenoménicas' y a la dificultad para encontrar razones para creer en su existencia.

otro camino al mismo resultado: lo mental no puede ser explicativo debido a su ineficacia causal.[25]

Sabatés considera, no obstante, que hay una respuesta plausible a ambos argumentos. El primer argumento descansa en el supuesto de que toda explicación es causal. Pero este supuesto, en su opinión, es demasiado estrecho. Su aceptación implicaría desechar explicaciones de propiedades sobrevinientes en términos de su base de superveniencia. Una opción más natural es aceptar un pluralismo explicativo que admita diferentes relaciones de dependencia como 'relaciones objetivas' que sirvan de base a la explicación. Sin embargo, advierte, sería apresurado afirmar que cualquier relación de dependencia produce una explicación apropiada. En el caso de la causación mental, el problema radicaría en el hecho de que la 'flecha de dependencia' que va desde un estado mental sobreviniente a su base tendría una dirección incorrecta. Si bien muchos tipos de dependencia, sostiene Sabatés, pueden ser explicativos, las *conversas* de esas relaciones de dependencia son típicamente no explicativas. Si un epifenomenista, concluye, desea mantener las explicaciones psicológicas, tiene que mostrar *por qué* la conversa de la superveniencia puede explicar un efecto de una base de superveniencia, mientras no puede explicar tal base, y también mostrar *por qué* la conversa de la superveniencia puede ser explicativa mientras que otras relaciones de dependencia (como la causación o la mera dependencia conceptual) no pueden serlo.[26]

Existe también una respuesta al segundo argumento. Si el argumento de la exclusión es correcto, una conducta debe tener una causa neurofisiológica. Esta causa neurofisiológica es necesaria [*necessitates*] para los estados intencionales que son citados en la explicación racionalizadora, y constituye su base de superveniencia. Esa causa neurofisiológica no necesita otras razones potenciales que el agente pueda tener para ejecutar la acción. En este modelo, se individualiza la razón por la cual el agente actuó como lo hizo y se descartan las otras razones para la acción que no están relacionadas a la causa real de la acción. El hecho de que podamos ser ignorantes acerca de lo que ocurre en el nivel neurofisiológico, concluye Sabatés, no

[25] Es de alguna manera irónico que desde el mismo ámbito filosófico en el cual surgieron las defensas de la tesis de que las explicaciones psicológicas *pueden* ser causales (Davidson, 1963, Fodor, 1968), surgieran dos décadas más tarde argumentos que parecen mostrar exactamente lo contrario.

debería molestarnos, dado que los estados intencionales nos proporcionan suficiente base para creer que una propiedad neurofisiológica apropiada está llevando a cabo el trabajo causal.

3. Estrategias ante el problema de la exclusión

En el mismo artículo, Sabatés realiza una amplia (pero no exhaustiva, según sus propias palabras) revisión de las alternativas de solución al problema. Dos grandes categorías de respuestas, en su opinión, pueden distinguirse: las estrategias incompatibilistas y las compatibilistas, cada una de ellas con variantes bastante dispares. Las primeras se caracterizan por negar uno de los dos supuestos que originan el problema: el materialismo no reduccionista o la causación mental. Las segundas se caracterizan por considerar que existe algún error en el argumento que conduce a la exclusión, y de que puede mantenerse tanto el materialismo no reduccionista cuanto la causación mental. Este tipo de estrategias se encuentra en la posición de tener que explicar que es lo que no ha sido entendido correctamente al plantear el problema de la exclusión.

I. Estrategias incompatibilistas:

1. El incompatibilismo contrario a la ortodoxia, variante que incluye :
   1.1. Un dualismo de sustancias al estilo cartesiano, que niegue el monismo. Esta posición, obviamente, debe enfrentar al problema de la interacción mente-cuerpo;
   1.2. Un dualismo fuerte de propiedades, que rechace la primacía de las propiedades físicas sobre las propiedades mentales, pero sin renunciar al monismo de sustancias. Esta posición debe enfrentar dos problemas: en primer lugar, la relación entre lo físico y lo mental resulta un completo misterio; en segundo lugar, no parece haber ventajas en el hecho de rechazar sustancias inmateriales no físicas si se sostiene a la vez que las propiedades mentales son absolutamente autónomas con respecto a las propiedades físicas.

[26] En el capítulo IV examinaremos con más detalle estas posibilidades, conjuntamente con la adopción de una posición que admita la ineficacia causal de la mente.

1.3. El emergentismo (interpretado a veces como una estrategia compatibilista), comprometido a la vez con alguna clase de tesis de dependencia que conecta las propiedades emergentes con las propiedades de base en distintos estratos, y con la causación descendente (de lo mental a lo físico, en el caso de propiedades de estas clases). Debido a esta última tesis, el emergentismo debe negar la clausura causal de lo físico y enfrentar el riesgo de caer en un dualismo de tipo cartesiano.

1.4. La negación del realismo de lo mental, en alguna de dos subvariantes:

1.4.1. una forma moderada que admite mantener las expresiones mentales por razones pragmáticas (irrealismo retentivo);

1.4.2. una forma fuerte que recomienda la eliminación del vocabulario mental (eliminativismo).

Ambas formas se caracterizan por considerar que el dualismo de propiedades es falso, ya que las propiedades mentales no son genuinas o reales. Dentro de esta subclase pueden incluirse las teorías de la identidad de tipos como una forma de irrealismo retentivo (ya que las propiedades mentales no son más que propiedades físicas), y, más claramente, la propuesta de Kim de las 'propiedades de segundo orden'. Todas estas posiciones disuelven el problema de la exclusión, ya que no hay causación mental de la cual ocuparse, pero tienen como consecuencia la pérdida de toda empresa científica que se ocupe de los fenómenos mentales, así como del conocimiento de sentido común sobre lo mental.

2. Incompatibilismo epifenomenista, que renuncia a la causación mental pero mantiene las restantes afirmaciones básicas del fisicalismo no reduccionista. El epifenomenismo es considerado en general inaceptable, ya que implica 'disolver' el problema de la exclusión simplemente renunciando a una de las más arraigadas intuiciones acerca de lo mental, que es su poder causal. El modelo de la causación superveniente de Kim y el de la explicación por programa propuesto por Jackson y Pettit son exponentes recientes de posiciones epifenomenistas.

El epifenomenismo, pese a su mala fama, merece, en opinión de Sabatés, ser explorado en detalle. Para que este enfoque pueda resultar plausible, dos problemas deben ser resueltos: la realidad y el poder explicativo de las propiedades mentales.

II. Estrategias compatibilistas:

1. El emergentismo es un primer candidato obvio, al afirmar que el fisicalismo necesita ser reinterpretado. Como se observó (1.3.), el emergentismo implica la negación de la tesis de la clausura causal del mundo físico. Las cuestiones que el emergentismo necesita afrontar son dos: en primer lugar, si el abandono parcial del carácter básico del mundo físico es suficiente para evitar las conclusiones del argumento de la exclusión; en segundo lugar, si una teoría que renuncia a la clausura causal del mundo físico puede ser considerada fisicalista en algún sentido.
2. Un segundo enfoque es el llamado de la 'primacía explicativa'. Lo mental, de acuerdo con este punto de vista, es causalmente relevante en la medida en que tenga lugar en explicaciones exitosas. Este enfoque, según Sabatés, desemboca en una concepción irrealista acerca de lo mental, tal vez de tipo retentivo.[27] Los predicados mentales resultarían meras herramientas explicativas y predictivas, y la realidad de las propiedades mentales no estaría entre las tesis a defender, a menos que se estuviese dispuesto a sostener una afirmación del tipo 'una propiedad es real si y sólo si figura en explicaciones exitosas'. Este tipo de afirmación, observa Sabatés, no refleja el punto de vista que el realista acerca de las propiedades mentales desea sostener.
3. El tercer y último enfoque dentro de este grupo de estrategias propone la reinterpretación de la causación mental en términos de una concepción contrafáctica de la causalidad. Un problema general que enfrenta esta alternativa, observa Sabatés, es que el testeo de la causalidad por medio de los contrafácticos es un test pobre para evaluar direccionalidad causal (y relaciones de dependencia en general).

[27] *Cfr.* el capítulo V para un análisis de esta estrategia.

A continuación examinaremos el rol que la investigación fáctica puede jugar en relación con los argumentos de la exclusión.

4. La pertinencia de la evidencia experimental para el análisis del problema

Muchas de las estrategias que acabamos de describir no apelan especialmente a los resultados de la investigación científica para apoyar sus tesis. Sin embargo, dado que no pocos autores conciben a la empresa filosófica como estrechamente integrada al campo de las ciencias (para algunos de ellos, al de la ciencia cognitiva), parece pertinente plantear el rol que juega –o debería jugar- la información científica relevante para el tratamiento del problema.

La pertinencia de la evidencia empírica para la evaluación de las doctrinas sobre el problema conduce a cuestiones espinosas, que llevan con frecuencia a la discusión acerca de la naturaleza de la reflexión filosófica y su relación con la investigación científica, muy cercanamente vinculada con los intentos de naturalización de distintos problemas filosóficos llevados a cabo en las últimas décadas. Dado que esta discusión trasciende ampliamente los límites de esta tesis, nos limitaremos a fijar una posición moderada al respecto: los datos provenientes de las ciencias (y, eventualmente, del sentido común), pueden ser pertinentes para evaluar tesis filosóficas y para dar razones en favor de su aceptación o rechazo; no son suficientes, por sí mismos, para refutar o verificar tales tesis. Hay varias razones para apoyar esta posición.

En primer lugar, la evidencia empírica no llega a nosotros interpretada. Los datos experimentales y observacionales son a menudo objeto de interpretaciones divergentes; y estas interpretaciones divergentes suelen ser incompatibles en sus presupuestos teóricos. Como consecuencia, muchas veces no resulta sencillo determinar cuando la información fáctica constituye un elemento contundente en contra de una determinada tesis filosófica. Como ejemplo de tal situación, analizaremos brevemente un caso muy estrechamente relacionado con el problema que nos ocupa: la relevancia de la evidencia empírica proveniente de las investigaciones experimentales de B. Libet. Previo a este análisis, mostraremos las dificultades que se presentan al intentar mostrar a partir de la evidencia empírica la

inadecuación de una tesis filosófica: el caso del argumento de la realizabilidad variable elevado en contra de la Teoría de la Identidad.

4.1. El argumento de la realizabilidad variable

El argumento de la realizabilidad variable, junto con una defensa de la concepción funcionalista de los estados mentales, fue expuesto por primera vez por Putnam en la década del '60 (en especial en su (1967)). El peso de este argumento en la pérdida de popularidad de la Teoría de la Identidad, según algunos autores,[28] ha sido decisivo. Sin embargo, la fuerza de este argumento y sus consecuencias han sido objeto de interpretaciones dispares.[29]

Rabossi (1995a y 1995b, especialmente en este último) analiza las implicaciones de este argumento. Al decir de este autor, la matriz teórica de la Teoría de la Identidad incluye entre sus tesis que los enunciados que afirman la identidad de los estados mentales y cerebrales expresan verdades contingentes. Tales enunciados son del mismo tipo que las verdades contingentes y *a posteriori* expresadas en afirmaciones tales como 'el calor es energía cinética media'. El desarrollo de la neurología validaría la identidad de los distintos fenómenos mentales tipo con los estados cerebrales que les correspondieran. El carácter contingente de la identidad tiene entre sus posibles consecuencias la posibilidad de que la neurofisiología demuestre que la teoría es científicamente inviable, que haya fenómenos mentales que no tengan una contrapartida fisiológica, y que haya estados neurofisiológicos que no sean correlacionables con estados mentales. Los primeros teóricos de la identidad, entonces, previeron la posibilidad de que su propuesta teórica fuese modificable (o aun desechable) a través de la evidencia empírica.

Esta última posibilidad dio lugar a la existencia de conocidos argumentos tendientes a refutar la Teoría de la Identidad, los cuales, pese a su éxito, no están exentos de críticas y contraargumentaciones.

---

[28] *Cfr*. Kim (1998), capítulo 1.

[29] Kim (1989b), sugiere una línea de crítica en contra del argumento. Este argumento, observa, 'quizás muestre que la conexión fuerte de las propiedades mentales *vis-à-vis* las propiedades físicas no tiene lugar; sin embargo, *presupone que la conexión fuerte específica por especies* si se da. Para desechar el argumento antirreduccionista, no necesito adscribir a esta segunda afirmación; todo lo que necesito es la afirmación más débil de que el fenómeno de la realización múltiple es *consistente* con la conexión fuerte específica por especies, y me parece que esto es claramente verdadero' (p. 274. Cursivas del autor)

El argumento de la realizabilidad variable, en opinión de Rabossi, puede ser concebido de dos maneras. Una de ellas es una lectura fáctica: el argumento partiría del hecho de la realizabilidad variable y permitiría concluir la imposibilidad factual de identificar tipos psicológicos con tipos neurofisiológicos (o las relaciones nomológicas que se dan entre ellos); la segunda lectura es una lectura conceptual: da por sentada la conclusión anterior y posibilita concluir la imposibilidad conceptual de la identidad. En opinión de este autor, la versión fáctica del argumento es 'sumamente cuestionable' por varias razones. Nos interesa particularmente una de ellas:

> El Argumento proclama el colapso de todo intento de correlacionar predicados psicológicos con estados físico-químicos únicos si se encontrara que dos organismos, O_1 y O_2, de los que se predica que están en un mismo estado psicológico P, tienen estados físicoquímicos diferentes. Ello supone que las semejanzas o diferencias de tales estados se pueden determinar con prescindencia de los esquemas clasificatorios pertinentes. Y esa suposición es errónea. Los criterios de semejanza o de diferencia entre estados son siempre relativos a tales esquemas. Si ése es el caso, resulta legítimo afirmar que O_1 y O_2 se encuentran en un mismo estado físicoquímico (con respecto a un cierto esquema teórico) y afirmar, al mismo tiempo y sin contradicción, que los sistemas nerviosos de O_1 y O_2 difieren en composición material o en organización (de acuerdo a otro esquema clasificatorio)' (1995b, p. 174).

Si admitimos que la clasificación de los estados (sean estos físicoquímicos, mentales, o de otra clase) depende de nuestros esquemas teóricos, y que no hay una forma neutral de clasificar tales estados, muchos de los intentos de emplear evidencia empírica para *refutar* tesis relativas al problema mente-cuerpo parecen de antemano condenados al fracaso.

Una argumentación semejante puede hacerse con respecto a los intentos de probar empíricamente la prioridad de los estados físicos con respecto a los estados mentales, o viceversa: tal posibilidad dependerá de los criterios de individuación y clasificación de los sucesos, y esos criterios dependerán de argumentos filosóficos tan controvertidos como cualquier otra tesis dentro de esa disciplina.

4.2. La prioridad de los sucesos cerebrales por sobre los sucesos mentales

Los experimentos de B. Libet, llevados a cabo a partir de la década del '60, han producido información empírica, la cual ha sido interpretada en ocasiones como

evidencia que prueba que los estados físicos preceden a los estados mentales. Esta evidencia ha sido discutida en relación con conocidas teorías acerca de lo mental, como la teoría de la identidad y algunas formas de dualismo interaccionista. Es innecesario observar que tales hallazgos han sido objeto de intenso debate y de interpretaciones divergentes.[30]

Grant y Gillet (1995), analizando de las investigaciones de Libet, parten del supuesto de que existe una correspondencia uno a uno entre estados identificables del cerebro y un rango de adscripciones mentales determinadas. Sólo a partir de esta base puede resultar de interés la pregunta acerca de la prioridad de miembros de un conjunto de sucesos sobre su contraparte para propósitos explicativos y de ontología.

Describen el propósito de estas investigaciones de la siguiente forma. Se trata de 'relacionar sucesos mentales con sus concomitantes físicos; o, más específicamente, la primera percatación conciente [*awareness*] de realizar un movimiento con la primera indicación detectable de éste en el cerebro, *i. e.*, la 'disposición potencial' [*readiness potential*] y otros sucesos fisiológicos que preceden al movimiento eventual' (p. 334). La estrategia utilizada para determinar el momento de la ocurrencia de los sucesos mentales se basó en tomar como punto de referencia la ocurrencia de otro suceso: la percepción de un punto móvil en una posición particular sobre una pantalla. La percepción del punto en una posición particular fue informada por el sujeto como simultánea con la percepción conciente de la intención. Este supuesto 'momento del suceso mental' que precede al movimiento voluntario fue relacionado con los registros electroencefalográficos que detectaron la 'disposición potencial', un suceso eléctrico en el cerebro que precede al movimiento voluntario. Empleando esta estrategia, Libet encontró que el tiempo de la toma de conocimiento conciente inicial, tal como es informada por el sujeto, tiene lugar alrededor de 300 a 500 milisegundos más tarde que la disposición potencial.

[30] Honderich (1984) observa que los resultados de las investigaciones de Libet han sido utilizados para argumentar en contra de ciertas doctrinas acerca de la relación entre lo mental y lo físico (como la Teoría de la Identidad y la teoría de la correlación psicofísica legaliforme) y en favor de otras (como el dualismo en la versión de Popper y Eccles); asimismo, podría ser utilizada para el debate libertad-determinismo, en favor de la doctrina del poder de originación. Sin embargo, considera que la interpretación de la evidencia proveniente de los experimentos da lugar a dos hipótesis distintas, y que la hipótesis preferida por Libet y sus colaboradores no es sostenida por la evidencia.

Estos hallazgos dieron lugar al siguiente argumento, denominado por Grant y Gillet 'argumento de la prioridad física':

> '(i) El suceso mental consistente en el intento de actuar causa el acto
> (ii) El suceso físico precede al suceso mental de intentar actuar, y es inconciente; por lo tanto
> (iii) El suceso físico *causa* y es explicativamente anterior al suceso mental
> (iv) Por lo tanto, los sucesos concientes son epifenoménicos en la comprensión de la conducta humana' (pp. 334-5).

Cualquier intento de establecer, vía estrategias experimentales, una relación precisa entre sucesos mentales y sucesos cerebrales asociados depende esencialmente de tres supuestos: a) cualquier experiencia conciente es potencialmente informable [*reportable*] sobre la base de la introspección; b) que el único medio de determinar la ocurrencia de las experiencias concientes son esos informes; c) que la experiencia conciente informable inmediatamente previa a la acción es identificada como la causa mental de la acción.

Sin embargo, contra a) y b), Grant y Gillet sostienen que no hay razones para sostener que todos los sucesos concientes son necesariamente informables. Por una parte, observan, no es difícil percibir introspectivamente que nuestras experiencias concientes tienen matices que no están representados completamente en lo que podemos informar a otros acerca de ellas. Aun aceptando que la experiencia conciente es esencialmente discursiva y por lo tanto informable, debería mostrarse que el momento del informe tiene un ajuste preciso con el momento del suceso que consiste en la intención de actuar. Hacer esto requeriría algunos criterios para determinar las características temporales de la intención y superar los desafíos al modelo de acción que se suponga.

Otro problema se presenta con la premisa (ii) del argumento. Si bien se solicitó a los sujetos del experimento que informaran la más temprana aparición de la percatación conciente que precede al movimiento (denominada A), no puede tenerse la certeza de que no haya tenido lugar cierta actividad mental preparatoria inmediatamente antes de que el sujeto 'decidiera' o formara la intención de (A). Resulta difícil, observan, descartar la posibilidad de sucesos mentales preparatorios que conducen al momento en el cual el sujeto señala que posee una intención

definida. De hecho, observan, en caso de ausencia de tal actividad preparatoria, la ocurrencia de la 'decisión' en sí misma sería completamente aleatoria, en el sentido de que el sujeto no podría anticipar que tomaría una decisión en ese momento. Parece difícil sostener, prosiguen, la idea de que tal acto mental puramente arbitrario o aleatorio es lo que se quiere decir con 'acto significativamente voluntario', especialmente en el contexto del experimento, en el cual la mente de los sujetos parecería constantemente ocupada con el requisito de decidir una intención de A. Puede afirmarse entonces, observan, que aun si tienen lugar sucesos cerebrales antes de que el sujeto forme la intención de A, esto no significa que los preliminares mentales de A sean precedidos por preliminares cerebrales; meramente implica que la intención manifestable de A no surge *de novo* sin ninguna preparación mental/cerebral.

En síntesis, la relación entre sucesos mentales y la información acerca de ellos parece 'considerablemente más tenue' que lo que Libet supone. Por lo tanto, resulta cuestionable que los informes de 'intención' hechos por los sujetos de Libet tengan alguna significación.[31]

5. Niveles explicativos en psicología cognitiva y autonomía epistemológica

5.1. El modelo de Marr

El surgimiento y desarrollo de la ciencia cognitiva ha mostrado, como rasgo epistemológico original, la presencia de la reflexión filosófica como parte integrante de una empresa intelectual presentada casi unánimemente como interdisciplinaria (Gardner, 1985). En efecto, suelen mencionarse diversas ramas de la filosofía (como la filosofía de la mente y la gnoseología), como implicadas en el desarrollo de este programa de investigación. Una combinación particularmente destacada de investigación empírica y análisis metateórico desarrollada dentro de este marco es provista por la obra de David Marr.

Conjuntamente con sus investigaciones sobre la percepción visual, Marr desarrolló una influyente y ampliamente comentada teorización acerca de los niveles

[31] Grant y Gillet sugieren otra línea de crítica del argumento en favor de la prioridad física, basados en lo que consideran debilidades metodológicas del diseño experimental. No las mencionaremos aquí, ya que

de análisis/explicación/descripción que deben distinguirse en las investigaciones en psicología cognitiva e inteligencia artificial. La teorización que incluye la distinción de tres niveles explicativos de Marr ha sido objeto de múltiples análisis en contextos conceptuales muy diferentes: desde el examen de su posible aplicación en el problema de la conciencia (Dennet, 1991), hasta la discusión acerca del individualismo y las explicaciones psicológicas (Burge, 1986), pasando por análisis acerca de las teorías de la arquitectura de lo mental (Ezquerro, 1995), e incluso en análisis filosóficos de la tecnología (Broncano, 2000), por mencionar sólo unas pocas temáticas en las cuales esta distinción ha sido analizada.[32]

En su obra principal y más conocida, *La visión* (1982), Marr dedicó una reflexión específica sobre cómo debía entenderse una explicación correcta de un sistema de procesamiento de información, como es el caso de la visión.[33] En su opinión, resultaba necesario distinguir tres niveles que, conjuntamente, permitirían comprender de manera completa un mecanismo de procesamiento de información:

1. El de una teoría computacional o de cálculo, o *nivel computacional*, en el cual se determina qué es lo que el sistema hace, y por qué lo hace;
2. El de la representación y el algoritmo, o *nivel algorítmico*, en el cual se elige una representación para la entrada y la salida de la información y el algoritmo que se utiliza para transformar una en la otra.
3. El de la implementación en el soporte físico, o *nivel implementacional*, esto es, los detalles sobre el modo en el cual se realizan físicamente el algoritmo y la representación.

---

nuestro propósito es solamente enfatizar la dependencia de la interpretación de los resultados del experimento de ciertos esquemas conceptuales cuestionables.

[32] Kim (1998) hace referencia a la obra de Marr en el contexto de la discusión de las concepciones estratificadas del mundo y la distinción entre niveles de análisis, explicación u organización.

[33] Hay referencias anteriores a la distinción, entre ellas: 'La inteligencia artificial es el estudio de problemas complejos de procesamiento de información que a menudo tienen sus raíces en algún aspecto del procesamiento biológico de información. El objetivo de esta disciplina es identificar problemas interesantes y resolubles del procesamiento de información para solucionarlos. La solución de un problema de esta índole se divide naturalmente en dos partes. En la primera se caracteriza la naturaleza implícita de un cómputo particular y se comprende su fundamento físico. Se puede considerar esta parte como la formulación abstracta de *qué* se computa y *por qué,* y me referiré a ella como la 'teoría' de un cómputo. La segunda parte consiste en algoritmos particulares para llevar a cabo un cómputo, así que especifica *cómo se* computa. La elección del algoritmo depende por lo general del *hardware* en que se llevará a cabo el proceso y pueden ser muchos los algoritmos que realicen un mismo cómputo. Por otra parte, la teoría del cómputo sólo depende de la naturaleza del problema del cual es solución' (Marr 1977, p.153).

Estas distinciones pueden ser mejor comprendidas si se las aplica a algún artefacto sencillo, y es lo que Marr hace al elegir como ejemplo a una registradora: en el nivel computacional encontramos que efectúa operaciones aritméticas, por lo cual la primera tarea es dominar la teoría de la adición. En el nivel algorítmico, pueden elegirse números árabes para las representaciones, y para el algoritmo pueden seguirse reglas como sumar determinados dígitos primero y 'llevar' dígitos si la suma excede nueve. En el nivel implementacional, los símbolos y los procesos pueden ser físicamente implementados por medio de un sistema de ruedas metálicas, o por medio de estados eléctricos de conjuntos de circuitos digitales.

Marr no menciona el tipo de relación que vincula a los sucesos de cada nivel; se limita a observar que los niveles están relacionados de modo lógico y causal.[34] Otra observación importante es que cada nivel tiene su lugar en la comprensión del proceso de procesamiento de la información. Sin embargo, también señala que 'como [los niveles] están relacionados sólo de un modo laxo, algunos fenómenos únicamente podrán ser explicados a uno o dos niveles' (p. 34). Así, por ejemplo, la neuroanatomía está vinculada principalmente al nivel de la realización física del cálculo; lo mismo ocurre con la neurofisiología. La psicofísica, por su parte, está relacionada de modo más directo con el segundo nivel, el del algoritmo y la representación.

La distinción de niveles explicativos propuesta por Marr, pese a su importancia e influencia, no fue la única propuesta de distinción de niveles explicativos que era necesario distinguir. McClamrock (1991) señala que una distinción similar se encuentra en la obra de Pylyshyn. Este autor distingue el nivel biológico (o físico), el nivel simbólico (o sintáctico o funcional) y el nivel semántico o intencional, equivalentes respectivamente a los niveles implementacional, algorítmico y computacional de Marr.

Horgan y Tienson (1993) señalan que para muchos filósofos y científicos cognitivos la relación entre los niveles medio y de base es la de *realización*. Consideran que el término empleado por Marr, 'implementación', con el cual se

[34] La idea de que la relación entre los niveles *es causal* (además de lógica) no parece compatible con la idea de que la relación entre los niveles es la de realización física. La relación de realización física es usualmente concebida como una relación sincrónica, a diferencia de la relación causal. Posiblemente se trate de un uso algo descuidado de la noción de causalidad.

hace referencia al nivel de base, es esencialmente el término empleado por los científicos computacionales para lo que los filósofos llamarían realización.[35] La relación entre el nivel medio y el nivel superior también puede ser considerada como una relación de realización. Horgan y Tienson argumentan este punto de la siguiente manera: dado que una función individual computable puede generalmente ser computada por una variedad de distintos algoritmos, algunos de los cuales emplearán diferentes representaciones que otros, habrá una relación uno-muchos entre a) la función de transición cognitiva en el nivel superior de Marr, y b) los algoritmos (con sus representaciones asociadas) que computan esa función. Pero debido a que los estados-tipo del nivel medio son individualizados funcionalmente, en términos de los algoritmos específicos en los cuales figuran, distintos algoritmos producen distintos estados-tipo en el nivel medio. Por lo tanto, concluyen, los estados mentales intencionales postulados por la teoría de la computación son múltiplemente realizables en el nivel medio por diferentes estados-tipo computacionales.

La relación de realización presente en el modelo de Marr implica además la idea de que múltiples estructuras distintas del nivel inferior pueden realizar procesos de nivel superior, esto es, la conocida tesis de la 'realizabilidad múltiple'. Esta tesis, anticipada por Putnam en un célebre artículo en la década del '60 y ampliamente admitida en el ámbito de la discusión acerca de la relación mente-cerebro, constituyó como hemos visto un argumento considerado muchas veces decisivo en contra de la posibilidad de reducción de los hechos psicológicos a los hechos neurofisiológicos.

Posiblemente inspirados en la tesis de la realizabilidad múltiple, algunos investigadores, basados en la propuesta de Marr, defendieron lo que podría llamarse una 'autonomía metodológica': la implementación computacional o física, en principio, no aportaría información esencial para la descripción de la capacidad del sistema, por lo cual no habría que ocuparse de ella; o bien, en lo que parece una formulación alternativa, algunos defendieron una 'autonomía epistémica': las teorías

[35] No obstante, algunos filósofos han considerado la relación entre estos dos niveles como una relación de identidad de tipos. No es obvio, por lo tanto, que el esquema de niveles explicativos sea vulnerable a alguna clase de argumento de exclusión. En caso de que la relación entre el nivel de base y el nivel superior fuese de identidad de tipos el argumento de exclusión no podría plantearse.

psicológicas sólo requieren de evidencias psicológicas.[36] Sin embargo, al decir de Gardner, Marr era un cognitivista cabal: consideraba que ninguna disciplina, por sí sola, era capaz de desentrañar procesos tan complejos como los que él estudiaba.

Sería erróneo afirmar, por otra parte, que el éxito de esta distinción ha sido un indicador de aceptación unánime entre los teóricos.

Pese a su ubicuidad y a su éxito teórico, parece plausible pensar que este modelo puede enfrentarse a algunos de los problemas derivados de los argumentos de la exclusión, como ocurre con otras teorías acerca de lo mental.[37]

Una primera posibilidad es examinar el esquema de niveles a la luz de la potencial competencia explicativa de las explicaciones provistas por las teorías elaboradas en los distintos niveles; esto es, un planteo de exclusión explicativa. Esta posibilidad tiene lugar únicamente en el caso de que las teorías de cada nivel provean de explicaciones completas e independientes. Si, por el contrario (y como parece estar presente en la observación de Gardner) una explicación completa de un

[36] Sterelny, *The Representational Theory of Mind* (1990), citado en Skidelsky (1996).

[37] Horgan y Tienson (1993), sostienen que el modelo de tres niveles de Marr puede ser generalizado 'produciendo una tipología tripartita de niveles que permanece neutral acerca de los supuestos claves del clasicismo existentes en la formulación original de Marr' (p. 159). Esta tipología genérica puede ser articulada, en opinión de estos autores, con distintos enfoques de lo mental, los cuales difieren del clasicismo y que pueden ser integrados en el contexto del marco conexionista en ciencia cognitiva. Esta posibilidad de combinar el enfoque de niveles de explicación u organización con el enfoque conexionista sugiere la posibilidad de que alguna clase de argumento de exclusión sea aplicable *también* para este último enfoque. El conexionismo no se ha mostrado en el pasado invulnerable a argumentos que lo comprometen con alguna clase de eliminativismo. Un conocido argumento de Ramsey, Stich y Garon explota esta posibilidad. Según Green (1997), el argumento puede ser expuesto de la siguiente manera. En los modelos conexionistas, a diferencia de los simbólicos, los sucesos físicos individuales no representan nada en absoluto; la representación está distribuida por sobre toda la actividad de la red. Si un modelo conexionista ejemplificara una proposición como 'Si A entonces B', no habría ninguna parte individual que representara a A, ninguna otra parte que representara a B, y ninguna tercera que representara la relación 'Si ... entonces...'. La proposición se representaría 'holísticamente' por la red entera. Esta característica de los modelos conexionistas da lugar al argumento RSG: dado que todas las creencias y deseos del modelo están almacenadas en forma superpuesta por la red entera, y porque la actividad de la red *entera* debe participar en cualquier salida conductual, de ninguna creencia o deseo, o pequeño subconjunto de creencias y deseos, puede decirse en ningún caso que es causalmente responsable para una salida dada; ellos son en cierto sentido conjuntamente responsables para cualquier rendimiento. Así, puede decirse que ningún conjunto finito de creencias o deseos explica conducta alguna: por ejemplo, Pedro no entró en el bar que porque quisiera una cerveza; sólo de la organización de su red cognoscitiva entera puede decirse que ha sido causalmente responsable. De este modo, la explicación de conducta sólo se encontrará en un análisis de la estructura de la red. Los autores de este argumento sostienen que las creencias y deseos, dado que no explican causalmente la conducta, simplemente deben eliminarse como otras estructuras hipotéticas que han resultado no tener ninguna función explicativa. Por lo tanto, concluyen, la adopción de un enfoque conexionista de la cognición compromete lógicamente a su adherente con el eliminativismo con respecto a las creencias y deseos. Green observa que el argumento no elimina que las creencias y deseos completamente; sólo los elimina de lo que se considera la cadena causal. Esta última observación no tiene una importancia menor: el eliminativista, como hemos visto en

sistema de procesamiento de información *debe* apelar a los tres niveles, entonces la exclusión explicativa no se plantearía.

Sin embargo, hay obvias características del enfoque de niveles de Marr que parecen ajustarse a los modelos que son vulnerables a los argumentos de la exclusión en una u otra variante. En particular, podemos destacar dos aspectos: en primer lugar, la idea de que la relación entre niveles es una relación de realización; en segundo lugar, que los estados de un nivel son caracterizados funcionalmente. Estos dos aspectos parecen bastar para aplicar al modelo de niveles el argumento propuesto por Block, el cual parece conducir del funcionalismo al epifenomenismo. En efecto, los estados de los niveles superiores (los del nivel de la teoría del cálculo y los del algoritmo y la representación), al ser definidos funcionalmente,[38] carecerían de poderes causales; en los casos standard (casos en los cuales la propiedad de segundo orden es definida en términos de un efecto, y tal efecto es producido sin reconocimiento alguno de la propiedad de segundo orden por un ser inteligente), los estados funcionales carecerían de efectos. Otra manera de presentar la dificultad surge si se considera la equivalencia entre los niveles distinguidos por Marr y los propuestos por Pylyshyn. Para explicar esta variante se hace necesario recurrir a la noción de propiedad causalmente relevante. Mientras que algunas propiedades de un suceso son causalmente relevantes para la producción de un efecto, otras no lo son. Un ejemplo clásico de tal posibilidad, debido a E. Sosa, es el siguiente: un disparo ruidoso y bien dirigido causa una herida mortal en la víctima; de las dos propiedades que hemos descripto del suceso (entre muchas otras posibles), una de ellas es causalmente relevante para la producción del suceso (el que el disparo sea bien dirigido), mientras que la otra (que el disparo sea ruidoso) no lo es. En su aplicación al problema de la causación mental la conclusión de este argumento es que los estados intencionales, o estados con contenido, no serían causalmente relevantes en virtud de su contenido, sino en virtud de sus propiedades puramente biológicas o físicas. Para expresarlo con una fórmula que ha devenido clásica, puede decirse que el cerebro es un procesador sintáctico, no semántico.

---

este capítulo, debe mostrar el nexo entre ineficacia causal, por una parte, e impotencia explicativa e irrealismo, por la otra.

El análisis de la propuesta teórica de Marr vuelve a poner de manifiesto importantes cuestiones relacionadas con la relación entre la reflexión filosófica y la investigación fáctica. ¿Cuál es el peso que la crítica filosófica debe tener sobre una propuesta programática que, según muchos autores, se ha revelado fructífera para guiar la investigación? ¿Hay diferencias relevantes en los casos en que los argumentos de la exclusión afecten a teorías filosóficas, como la de Davidson, o a los fundamentos de programas de investigación empírica, como el de Marr? ¿Deberían esta clase de programas juzgarse sólo por sus resultados (esto es, fundamentalmente por la capacidad explicativa y predictiva de las teorías resultantes)? ¿O la crítica filosófica puede constituir un factor decisivo a la hora de evaluar la viabilidad de esta clase de programas?[39] Hemos observado que todas estas cuestiones, vinculadas con la manera en que la filosofía se articula con la ciencia, exceden los alcances de este trabajo. Sin embargo, señalarlas no carece de importancia.

Posiciones extremas pueden hallarse tanto en uno cuanto en el otro extremo del espectro de posibilidades. Por una parte, podría objetarse que la crítica filosófica no es pertinente; podría argumentarse que los enfoques metateóricos deben ser juzgados por el éxito (en particular, explicativo y predictivo) de las teorías resultantes de su aplicación. Por la otra, podría sostenerse que la crítica filosófica es pertinente para criticar programas erróneos, por lo que sus prescripciones no pueden dejar de ser tomadas en cuenta a la hora de tomar decisiones metodológicas en ciencia fáctica.

La plausibilidad de estas posiciones depende de cómo se conciba la relación entre las disciplinas que integran la ciencia cognitiva: si se concibe a la filosofía simplemente como una disciplina cuyo objetivo en tal contexto es simplemente proveer de una suerte de 'legitimación' metateórica de las estrategias de

[38] *Cfr.* el capítulo I.

[39] Una respuesta afirmativa a esta pregunta parece hallarse en esta afirmación de Kim: 'Si las preocupaciones de Block están bien fundadas, los funcionalistas están en un serio problema: las propiedades mentales estarían amenazadas con la pérdida de sus poderes causales, y el hecho de que un suceso caiga bajo un determinado tipo mental (digamos, dolor) no tendría ninguna relación con los efectos que pueda causar. En síntesis, el funcionalismo podría resultar ser una forma de epifenomenismo, y la concepción recibida ('oficial') de la ciencia cognitiva como una ciencia especial autónoma, que genera sus propias explicaciones causales distintivas basadas en leyes en los niveles cognitivos superiores, formales/abstractos, enfrenta un inminente colapso' (pp. 51-52)

investigación empírica, entonces tales análisis serían, en el mejor de los casos, superfluos; en el peor, perjudiciales, ya que distraerían los esfuerzos de la comunidad científica en criticar los fundamentos de programas de investigación que deberían ser juzgados por las teorías resultantes de su aplicación, y no por su adecuación a análisis *a priori*. Si, por el contrario, se concibe a la filosofía como una disciplina que puede contribuir positivamente a la investigación empírica, ya sea en la crítica de supuestos epistemológicos como en la elucidación de conceptos clave, las posiciones extremas perderán buena parte de su sustento.[40] Como hemos mencionado, aceptamos al respecto un punto de vista moderado, que sostenga que los resultados de la investigación filosófica son pertinentes para evaluar la corrección de la investigación empírica, aunque claramente no suficientes para decidir al abandono de programas de investigación.[41]

En el siguiente capítulo formularemos una serie de consideraciones acerca de la medida en que los argumentos de la exclusión afectan a las explicaciones psicológicas. Algunos de los supuestos que mantendremos acerca de la relación entre la reflexión filosófica y la investigación fáctica serán expuestos en el capítulo IV.

---

[40] 'El dominio de la explicación psicológica es uno en el cual la imposición de restricciones «filosóficas» sobre la teoría puede ser especialmente crítico debido a la naturaleza nebulosa del tema. Desafortunadamente, es también un dominio en el cual aquellas restricciones ideológicas a menudo han desorientado más que ayudado –lo cual es atestiguado por el conductismo o el introspeccionismo-. Debido a esto, deberíamos estar particularmente atentos a la posibilidad de que se deslicen y contribuyan a la confusión restricciones no razonables y quizás inadvertidas –hay ya, después de todo, abundancia de ellas en circulación-' (McClamrock, 1991).

[41] Parece pertinente al respecto la siguiente observación de Newton-Smith: '[A] diferencia de Popper, la mayor parte de los filósofos de la ciencia tienen el grado debido de modestia. Atentos a los notorios fracasos de los filósofos cuando intentaron explicar a los físicos como tiene que ser el mundo (Kant sobre el carácter euclidiano del espacio, por ejemplo), son conscientes también de los peligros que acechan a la tentación de explicar a los físicos cómo deberían proceder a la hora de comparar los méritos de las teorías' (1981, p. 27).

CAPÍTULO III: LAS EXPLICACIONES PSICOLOGICAS

1. La multiplicidad de las explicaciones psicológicas

Hemos señalado en el capítulo anterior que los argumentos de la exclusión causal, *prima facie*, parecen afectar nuestras concepciones acerca de la explicación psicológica (si bien hemos coincidido con Sabatés en que sería un error suponer sin más la irrelevancia explicativa de lo mental). Como mínimo, obligaría a aceptar una posición 'reformativa' con respecto a una parte de las explicaciones psicológicas: si es verdad que muchas de esas explicaciones presuponen la eficacia causal de los estados mentales, su legitimidad debería ser seriamente reconsiderada. Y, en ocasiones, parece considerarse que la psicología (o partes importantes de ella), ve amenazada su propia existencia como disciplina autónoma como consecuencia de tales argumentos.[42] Sin embargo, dada la enorme cantidad y heterogeneidad de los *explananda* psicológicos, parece razonable suponer que no existe un solo tipo de explicaciones psicológicas, sino muchos. Y no hay razones para suponer, *a priori*, que todas las explicaciones pertinentes para esos fenómenos serán necesariamente causales,[43] y menos aún que se verán afectadas irremediablemente por los argumentos de la exclusión.[44] Por otra parte, es obvio que esto último no implica

[42] Recuérdese al respecto el comentario de Kim sobre el 'inminente colapso' que enfrentaría la ciencia cognitiva –y, con ella, la psicología cognitiva- (*supra*, p. 63, *n.* 39).

[43] Cummins (1983), en su conocido estudio sobre la explicación psicológica, sostiene que 'los fenómenos psicológicos no son explicados típicamente subsumiéndolos bajo leyes causales, sino tratándolos como manifestaciones de capacidades que son explicadas por medio del análisis' (p. 1). En su opinión, una de las más desafortunadas consecuencias del predominio del modelo nomológico deductivo de explicación ha sido el concentrar la atención sobre las leyes causales y las explicaciones correspondientes, descuidando el estudio de explicaciones alternativas más importantes. La defensa de la idea de que las explicaciones psicológicas son principalmente no causales aparece también en ocasiones en autores que proponen paradigmas 'alternativos' para la psicología: 'Al introducirse en un lenguaje cienticista (...), gran parte de la psicología moderna ha supuesto simplemente que la forma causal de explicación es la correcta' (Harré, Clarke y De Carlo (1985), p. 21). Según estos autores, para lograr explicaciones psicológicas realmente científicas se debe seguir el modelo proporcionado por las explicaciones de las ciencias biológicas.

[44] Parece razonable sostener que hay explicaciones psicológicas no causales de un mismo fenómeno que no se verían, *prima facie*, afectadas por los argumentos de la exclusión. Un interesante ejemplo de esta clase de explicaciones es proporcionado por la psicología del desarrollo en el área de estudio de los fenómenos de apego. Es un hecho bien confirmado la presencia, en niños mayores de seis meses, de reacciones de temor o ansiedad ante la presencia de desconocidos. Este fenómeno ha recibido explicaciones de diferentes tipos, existiendo considerables divergencias entre ellas. Una clase de explicaciones ha sido de corte cognitivo-social: se ha afirmado que la reacción se produce por la discrepancia entre las propiedades estimulares del desconocido y la representación de las figuras familiares para el niño; otra explicación ha apelado a la interrupción de las expectativas y el plan de

aceptar que todas las clases de explicaciones psicológicas que se han propuesto resulten epistémicamente satisfactorias, o que las teorías a partir de las cuales se proponen lo sean; una explicación psicológica puede no verse afectada por los argumentos de la exclusión pero ser, por diversas razones, inadecuada.

La pregunta relativa a cuáles son los criterios que permiten caracterizar a una explicación psicológica como adecuada o satisfactoria es tan problemática como la pregunta referente a qué es una explicación científica satisfactoria. Las últimas décadas atestiguan un intenso debate acerca de qué constituye una explicación científica adecuada. Por citar sólo los enfoques más conocidos, se desarrollaron en ellas desde los modelos de la explicación que podrían considerarse 'clásicos', que enfatizaron el carácter inferencial y nomológico de la explicación, propuestos por Hempel (1942, 1948, 1965a, 1966) y Popper (1934, 1972), hasta los enfoques de la pragmática de la explicación (Van Fraassen, 1980), pasando por concepciones de la explicación como unificación, indudables herederas de los primeros (Friedman, 1974), y por modelos de explicación mecánico-causal (Salmon, 1984). Esta controversia, aún vigente, hace pensar que el recurrir a la metateoría desarrollada por filósofos o por psicólogos indudablemente mejorará nuestra comprensión de los aspectos ontológicos y epistemológicos de la explicación psicológica, pero no parece probable que nos proporcione una respuesta concluyente a la cuestión planteada. Parecen resultar pertinentes, respecto de la multiplicidad de teorías sobre la

---

acción del niño. Sin embargo, otros autores han señalado que tales explicaciones, si bien contribuyen a identificar los mecanismos que tienen lugar en la reacción, no determinan el sentido funcional de la conducta. Según este enfoque, esta conducta de temor ante lo desconocido por el solo hecho de ser desconocido tendría la función de proteger al niño de los peligros en el momento en que sus capacidades de locomoción se desarrollan y comienza a separarse espacial y temporalmente de la madre. Podemos considerar que las explicaciones cognitivo-sociales y etológicas o funcionales no son incompatibles ni se encuentran en situación de competencia: ambas proveen de una comprensión de diferentes aspectos del fenómeno. Ahora bien, la segunda explicación no parece causal: las explicaciones funcionales (al menos para muchos autores) son tipos específicos de explicación, no reductibles a explicaciones causales. Recae en el defensor de la exclusión causal-explicativa el intento de mostrar cómo tales explicaciones son reductibles a explicaciones causales, y, en caso de que esto último pueda lograrse, mostrar cómo se ven afectadas por los argumentos de la exclusión. (O también, como analizaremos en el capítulo VIII, mostrar cómo podemos considerarlas explicaciones complementarias de un único *explanandum* poco definido). Pinker (1997) expresa una idea similar a la expuesta. Al exponer el programa de la psicología evolucionista, observa que 'la ciencia cognitiva nos ayuda a comprender cómo es posible la mente y de qué clase es la que tenemos. La biología evolutiva nos ayuda a entender *por qué* tenemos la clase de mente que tenemos' (p. 42. Cursivas del autor).

explicación, las observaciones de Kim acerca de la laxitud y flexibilidad de la noción de explicación científica.[45]

Es interesante observar, por otra parte, que aun en el caso de que los argumentos de la exclusión afectaran a sólo una parte de las explicaciones psicológicas esto no implicaría aceptar sin mayor análisis el éxito explicativo de la psicología. Parece difícil negar que ha habido progreso en la psicología contemporánea. Este progreso incluye varias dimensiones: el descubrimiento de nuevos hechos y regularidades desconocidos para la psicología de sentido común, la aplicación de los conocimientos básicos en el diseño de tecnologías que permiten conocer y modificar de manera fundada los hechos psicológicos, el desarrollo de praxiologías que permiten la acción racionalmente planificada, y, en algunos casos, el incremento de la capacidad para predecir conductas futuras. Sin embargo, resulta más difícil justificar la afirmación de que ha habido progreso explicativo.

La psicología contemporánea ha proporcionado explicaciones novedosas a fenómenos bien conocidos, a la vez que ha propuesto explicaciones para fenómenos nuevos. Ahora bien, la competencia explicativa es un fenómeno bien conocido en el ámbito de esta disciplina; esto es, la existencia de teorías explicativas alternativas, en ocasiones incompatibles, que explican rangos similares de fenómenos. Si bien en casos especiales la existencia de más de una explicación puede ser considerada aceptable,[46] la sobreabundancia sistemática y sostenida en el tiempo de varias explicaciones para los mismos fenómenos no puede ser considerada satisfactoria, ya que implica la existencia de teorías rivales, muchas de las cuales, al menos desde una perspectiva mínimamente realista de la ciencia, no pueden ser correctas.[47] La situación resulta menos satisfactoria aun si tenemos en cuenta que muchas de esas explicaciones pueden resultar cuestionables, ya sea por el modo en que se ha llegado

[45] *Infra*, p. 111.
[46] Véase al respecto el capítulo VIII.
[47] Un panorama más ajustado de la psicología actual en este respecto es indudablemente más complejo que la somera descripción expuesta. Muchos de los análisis histórico-epistemológicos del desarrollo de la psicología se plantean en términos de competencia de paradigmas, programas de investigación o alguna otra clase de unidad de análisis mayor y más compleja que la teoría. En este caso estarían implicadas no sólo divergencias teóricas, sino también metodológicas y metateóricas. Sin embargo, para poner de manifiesto la situación de competencia explicativa es suficiente con una descripción simplificada como la expuesta.

a postular los principios explicativos como por la insuficiente confirmación de la existencia de la misma regularidad que se desea explicar.

Dado que nuestro interés principal no es la explicación psicológica *simpliciter*, sino la relación entre el problema de la exclusión causal de las propiedades/estados mentales y la explicación psicológica, en este capítulo intentaremos mostrar, en primer lugar, cómo los argumentos de la exclusión impactan de manera diferencial en los distintos tipos de explicación psicológica, y que este impacto será muy variable dependiendo de la amplitud de los criterios empleados para admitir explicaciones como satisfactorias; en segundo lugar, examinaremos un intento de mostrar cómo ciertas explicaciones psicológicas no se ven afectadas por los argumentos de la exclusión. En el curso de este análisis podrá observarse que, si bien en algunos casos los supuestos causales en las explicaciones son evidentes, en otros están ocultos o implícitos.

2. Las explicaciones en psicología social

Es frecuente que la discusión relativa a la posibilidad de la explicación psicológica, aceptados los argumentos de la exclusión causal, se centre en el análisis de la manera en que la psicología cognitiva resultaría afectada.[48] Sin embargo, hay otras áreas de la psicología que se verían afectadas tan directamente (o más) que la psicología cognitiva. Una de ellas es el área de la psicología social. Muchas teorías en este campo presuponen una estructura conceptual formada por estados intencionales a los que se reconoce eficacia causal. Si los argumentos de la exclusión fuesen correctos, nos veríamos enfrentados, en el mejor de los casos, a la necesidad de adoptar una posición 'reformativa' con respecto a las explicaciones proporcionadas en este campo. Esto es, a reformular o reinterpretar tales explicaciones para adecuarlas a las conclusiones del argumento. Describiremos a continuación un ejemplo especialmente claro de esta posibilidad.

La teorización referente a las actitudes en el campo de la psicología social debió enfrentar el crónico problema de la consistencia entre actitudes y conducta. Si bien clásicamente se admitió que la actitud debería ser un predictor confiable de la

[48] El título del artículo de Sabatés (1996) es elocuente al respecto.

conducta, la investigación se encargó de desmentir reiteradamente este supuesto. Varios desarrollos teóricos intentaron superar esta dificultad.

Icek Azjen y Martin Fischbein (1980) desarrollaron durante la década del '70 la denominada 'Teoría de la Acción Razonada'. Este intento teórico incluyó como objetivos predecir y comprender las influencias motivacionales sobre la conducta que no están sometidas al control volitivo, identificar estrategias para el cambio de la conducta y explicar virtualmente cualquier acción humana, incluyendo acciones tales como comprar un auto nuevo, votar contra un candidato y ausentarse del trabajo.

La teoría de la acción razonada es una teoría general de la conducta humana que trata con la relación entre creencias, actitudes, intenciones y conducta. En términos generales, la teoría sostiene que las conductas son una función de las intenciones que existen para realizarlas; a su vez, estas intenciones se encuentran determinadas por actitudes hacia la realización del comportamiento y por normas subjetivas con respecto al mismo. Por último, esas actitudes y normas subjetivas están determinadas por creencias conductuales y creencias normativas respectivamente.

El paso inicial para la aplicación de la teoría en un área determinada es la identificación de conductas de interés. La identificación completa de una conducta requiere tomar en consideración cuatro elementos: acción, objeto, contexto y tiempo. Un cambio en cualquiera de estos cuatro elementos produce una redefinición en la conducta de interés.

Ya identificada la conducta, la teoría sostiene que el mejor predictor simple de esa conducta es la intención de la persona de realizarla. Además, se supone que la mayor parte de las acciones humanas socialmente relevantes son controladas voluntariamente, y, por lo tanto, la intención de realizar esas conductas es el principal determinante de las mismas.

En tercer lugar, la teoría sostiene que las intenciones de una persona son función de dos determinantes básicos, uno personal y otro social. El factor personal se denomina actitud hacia el comportamiento, y hace referencia a los sentimientos positivos o negativos de la persona con respecto a realizar la conducta en cuestión. El factor social se denomina norma subjetiva, y se refiere a la percepción de la persona acerca de las presiones sociales que lo llevan a realizar o no la conducta.

En líneas generales, los individuos intentarán realizar una conducta cuando tengan una actitud positiva hacia su realización y cuando crean que personas importantes para ellos piensan que tal conducta debería ser realizada. Por otra parte, si bien las actitudes y las normas subjetivas influyen en la formación de cualquier intención dada, la importancia relativa de ambos factores puede variar de acuerdo con cuál sea la conducta, el individuo, o aun la cultura de la cual se trate.

Si bien el nivel de explicación anterior ofrece un programa inicial de porqué las personas se comportan como lo hacen, una comprensión más completa de las intenciones requiere a su vez una explicación de por qué la gente mantiene unas actitudes o normas subjetivas determinadas. Con respecto a lo primero, la teoría sostiene que la actitud de una persona para realizar una conducta determinada se halla en función de la totalidad de sus creencias más 'importantes', las cuales le indican que la realización de la conducta le proporcionará ciertos resultados y una evaluación personal de esos resultados. Con respecto a lo segundo, se considera que la norma subjetiva de la persona con respecto a la realización de una conducta es función de las creencias normativas acerca de que ciertos grupos o individuos importantes consideran que esa misma persona debería realizar o no esa conducta.

Esta teoría, aun cuando no utilice lenguaje causalista, parece presuponer la relación causal:

> [P]artimos del supuesto de que muchas acciones de relevancia social se encuentran bajo control volitivo y, consistentemente con este supuesto, nuestra teoría concibe la *intención* de la persona de llevar a cabo (o no hacerlo) una conducta como el determinante inmediato de la acción (...) [L]a intención de una persona es una función de dos determinantes básicos, uno de naturaleza personal y el otro un reflejo de la influencia social (Ajzen y Fishbein, 1980, pp. 5-6. Cursivas de los autores).[49]

Las intenciones que determinan las conductas las preceden, entonces, en una relación temporal asimétrica (causación psicofísica); a su vez, las intenciones son precedidas y determinadas por actitudes y por normas subjetivas (causación de un estado mental por otros estados mentales). Si esta interpretación es correcta, no

[49] Podría argumentarse que, aun cuando se considere a la forma de determinación en juego como causal, los autores no se pronuncian acerca de la ontología de los estados mentales y sus relaciones. De esta forma, la idea de que los estados mentales podrían tener poderes causales sería compatible con alguna

cabe duda que argumentos como los de la exclusión causal impactan directamente en contra de explicaciones como éstas: ni las intenciones pueden ser causas de la conducta ni estados mentales tales como las actitudes y las creencias pueden ser causas de otros estados mentales, como las intenciones; las explicaciones que involucren tales estados, por lo tanto, no podrán ser interpretadas como causales.[50] Esta situación se repetirá, en principio, en muchos otros casos de teorías que apelen a tales estados intencionales.

Por lo tanto, si los argumentos de la exclusión son sólidos, tales explicaciones no pueden perdurar en su forma actual. Resultaría necesario reinterpretarlas, quizás a la manera de la explicación por programa propuesta por Jackson y Pettit.[51] Los estados mentales a los que se hace referencia en tales explicaciones podrían mantener su status de constructos útiles para la predicción de la conducta; sin embargo, no podrían ser considerados como causas de ella.

3. Una clasificación de las explicaciones en psicología

Nos referiremos ahora a una clasificación de las explicaciones psicológicas debida a Piaget (1963). Si bien nuestro interés se centra en la clasificación que Piaget realiza de las formas de explicación psicológica, tendremos previamente que describir el concepto piagetiano de explicación. Esta clasificación, pese a presentar varios aspectos cuestionables (que comentaremos luego), presenta el interés de describir tipos de explicación efectivamente utilizados en psicología –no meramente propuestos de manera programática o tentativa- y plantear, quizás involuntariamente, la cuestión del límite que separa las explicaciones psicológicas de las explicaciones extrapsicológicas de los fenómenos psicológicos.

Piaget postula tres clases de procedimientos que caracterizan la investigación psicológica destinada a la construcción de explicaciones:

---

forma de epifenomenismo de tipos. Sin embargo esta interpretación resultaría forzada y, nos parece, *ad hoc*.

[50] En caso de que se adopte una concepción de los sucesos como particulares complejos, un determinado estado no será capaz de causar algo *en virtud de* sus propiedades mentales.

[51] *Infra*, p. 107.

a. En primer lugar, el establecimiento de regularidades o leyes. En sí misma, la ley no explica nada, ya que se limita a enunciar la generalidad de una relación de hecho.
b. En segundo lugar, a la legalidad postulada se suma la construcción deductiva: la explicación supone un sistema de leyes tal que una de ellas pueda construirse o reconstruirse deductivamente a partir de las otras, primer rasgo éste que diferencia a la explicación de la mera legalidad. Sin embargo, la deducción de una ley a partir de un conjunto de otras leyes no constituye aún una explicación causal.
c. En tercer lugar, un procedimiento que complementa necesariamente a los anteriores y que constituye el segundo carácter específico de la explicación causal: la deducción de la ley que hay que explicar 'no es simplemente ideal o "lógica", sino que se aplica a un sustrato "real" o "modelo" que se supone se adapta a tal deducción y "representa" sus diversas relaciones' (p. 159).

Ahora bien, Piaget observa que, desgraciadamente, existe una multiplicidad de tipos de explicaciones posibles en psicología, muchos más que en biología, química o física. La razón principal de esta multiplicidad estriba, en su opinión, en la diversidad de lo que ha llamado los 'modelos' que sirven de sustrato a las relaciones deductivas que se han establecido entre las leyes. Y esta abundancia de modelos, sostiene Piaget, se debe esencialmente a las dificultades planteadas por la necesidad de dar una solución aceptable y fecunda al problema de las relaciones entre las estructuras de las reacciones concientes y las estructuras orgánicas. Por lo tanto, en alguna medida Piaget hace depender el problema de la multiplicidad de las explicaciones psicológicas de la solución al problema mente-cerebro.

Teniendo en cuenta que el criterio utilizado para la clasificación son los distintos modelos, Piaget considera que existen dos grandes tipos, o, como él mismo señala, dos polos en los modelos explicativos corrientes:

a. Los que se orientan hacia una reducción de lo más complejo a lo más simple o de lo psicológico a lo extrapsicológico, o
b. Los que se encaminan hacia alguna clase de constructivismo que se mantiene en mayor u menor medida dentro de los límites de la 'conducta'.

Según Piaget, dado que los modelos de tipo reduccionista pueden buscar la reducción principalmente en el ámbito de lo psicológico, o bien tender a la reducción de lo mental a realidades externas a este campo, habrá tres grandes categorías (A-C), y, dentro de las dos últimas, tres variedades:

A. Una forma de reduccionismo psicológico.
B. Diversas formas de reduccionismo extrapsicológico:
   B_1. Las explicaciones sociológicas.
   B_2. Las explicaciones fisicalistas.
   B_3. Las explicaciones organicistas.
C. Las explicaciones 'constructivistas'
   C_1. Los modelos del tipo 'teoría de la conducta'.
   C_2. Los modelos de tipo genético.
   C_3. Los modelos llamados 'abstractos'.

A. El primer tipo de explicación, a la que Piaget denomina reduccionismo psicológico, consiste en buscar la explicación de una serie de reacciones o conductas variadas por la reducción a un mismo principio causal que permanece inmodificable durante las transformaciones. El ejemplo que propone de este tipo de explicación es proporcionado por los trabajos experimentales de psicoanalistas de orientación freudiana acerca del desarrollo de las relaciones 'objetales'. En sus primeras formulaciones, la explicación psicoanalítica del desarrollo afectivo se basaba puramente en el concepto de *libido*, la energía psíquica en primer lugar concentrada en ciertas actividades orgánicas (etapas oral y anal), desplazándose luego al conjunto de la actividad propia (narcisismo) y, por último, a las personas exteriores (elección de 'objeto' y relaciones objetales). Las novedades, observa Piaget, son sólo el resultado de un desplazamiento de las cargas afectivas y no de una estructuración cognitiva. Trabajos posteriores, observa Piaget, moderan esta posición extrema y admiten el rol de esta última.

B_1. El primer tipo de explicación reduccionista intenta la reducción de lo psicológico a lo extrapsicológico, en este caso a lo social. Ejemplos de este tipo de explicación Piaget lo encuentra, entre otros casos, en el psicoanálisis denominado 'culturalista' (Fromm), en los trabajos de Vigotski y Luria sobre el lenguaje, y en la sociometría. Desde esta perspectiva, observa Piaget, cuando

una conducta nueva viene a enriquecer o ampliar las anteriores en el curso del desarrollo, se la considera más bien el aporte resultante de las interacciones de la vida social que el resultado de una construcción interna. No excluye la construcción, observa Piaget, pero la construcción se desplaza al terreno de las interacciones colectivas en vez de ser el resultado de mecanismos psicobiológicos.

B_2. El segundo tipo de explicación que intenta la reducción de lo psicológico a lo extrapsicológico busca la reducción fisicalista. El ejemplo que proporciona Piaget de esta posición está dado por las explicaciones propuestas por la teoría de la Gestalt, que no sólo tienden a reducir los hechos mentales a hechos fisiológicos, sino que también tratan de subordinarlos, mediante los esquemas de campo, a estructuras físicas, de lo cual surgen interpretaciones que conducen casi directamente de lo psicológico a lo físico.

B_3. El tercer tipo de explicación que intenta la reducción de lo psicológico a lo extrapsicológico busca la reducción organicista. Como para muchos autores, observa Piaget, el dominio de lo psicológico constituye la zona de interferencia entre lo biológico y lo social, el modo privilegiado de explicación reservado al psicólogo, en aquellos puntos en los que no esté subordinado a la sociología, será la reducción de lo superior a lo inferior, esto es, la asimilación a los modelos fisiológicos. Ejemplos de este tipo de explicaciones pueden hallarse en la obra de Pavlov.

Pasando ahora a los tipos de explicaciones denominados 'constructivistas', Piaget aclara que estos modelos explicativos no rechazan ninguna de las formas de explicación anteriores, pero las complementan con un constructivismo específicamente psicológico.

C_1. En primer lugar se encuentra lo que denomina 'la explicación por la conducta'. Dentro de esta clase Piaget incluye en especial a la teoría del aprendizaje de Hull (y también a la de Tolman), encuadradas dentro del conductismo. En este tipo de explicación, en particular haciendo referencia al sistema de Hull, la conducta es explicada por medio de ciertos conceptos centrales –en especial, el de 'familias jerárquicas de hábitos-, que tienen poder explicativo

independientemente de las posibles reducciones organicistas a las que no se recurre en este tipo de esquemas explicativos.

C_2. El segundo tipo de explicación constructivista es denominado por Piaget 'por construcción genética'. A diferencia del modelo anterior, propuesto por los teóricos del aprendizaje, en el cual el papel central para la explicación del desarrollo mental estaba jugado por el concepto de aprendizaje, combinándose con la maduración, para este modelo la maduración y el aprendizaje constituyen sólo dos de los factores en juego. Además, sustituirá la noción de comportamiento por la noción de 'conductas', que define como el comportamiento más las acciones interiorizadas que van acompañadas con 'toma de conciencia' (ya que no rechazan los conceptos mentalistas que están prohibidos para los defensores del modelo anterior). Como ejemplo de teoría que sostiene este tipo de modelo explicativo Piaget cita partes de su propia obra.

C_3. Como último tipo explicativo de la clase de explicaciones constructivistas encontramos la llamada 'explicación fundada en los modelos abstractos'. En base a las tres características que Piaget atribuye a la explicación, la explicación fundada en modelos abstractos puede definirse de dos maneras, una general y otra particular:

a. De modo general, Piaget observa que se recurre a modelos abstractos cuando, en vez de mantener, como esquema de la deducción, con la deducción 'ingenua' fundada en el lenguaje corriente, se adopta un esquema deductivo de carácter técnico, tomado de la matemática o la lógica. Según esta definición, la elección de un modelo abstracto no cambia el sustrato elegido en 3).
b. De manera particular, se hablará de una explicación por modelo abstracto cuando, para un conjunto de leyes, se utilice un esquema deductivo técnico 2), pero sin elegir un sustrato real determinado 3) e intentando sustituirlo por lo que puede haber en común entre los diferentes modelos posibles. El modelo es 'abstracto' en el sentido en que 'abstracto' es 'simplemente común a los diferentes modelos reales que puedan concebirse'.

La explicación por modelos abstractos, en su forma general o particular, brinda, según Piaget, tres clases de servicio:

a. Otorga precisión a las deducciones que en caso contrario serían imprecisas.
b. Permite descubrir relaciones nuevas entre hechos generales o leyes anteriormente no comparables.
c. Puede proporcionar nuevas relaciones causales que el análisis no había detectado.

Varias observaciones son pertinentes respecto de esta clasificación. Recordemos, en primer lugar, que el criterio de clasificación es la diferencia entre modelos y no características formales o lógicas de los tipos de explicación descriptos; esto es, básicamente el nivel del cual se extraigan las leyes explicativas. Se admite, por otra parte, que podrían existir tipos intermedios.

Dos aspectos cuestionables son el de considerar explicaciones psicológicas a estrategias reduccionistas, y el de considerar causales a explicaciones que visiblemente parecen no serlo. Es opinable que las estrategias B_1, B_2 y B_3, que como claramente señala Piaget plantean la reducción de lo psicológico a diversos principios y mecanismos extrapsicológicos, puedan ser consideradas explicaciones psicológicas. Por el contrario, de lo que parece tratarse es de extraer la explicación de los fenómenos en cuestión del campo de la psicología y situarla en disciplinas adyacentes. Esto, por supuesto, no implica invalidar su valor como estrategia de investigación, en tanto que cumplan con los objetivos y presenten las ventajas que usualmente se atribuyen a la reducción. Por otra parte, y en relación con el problema que nos ocupa, tales consideraciones no deberían inquietar en lo más mínimo al defensor de posiciones reduccionistas en psicología. Por el contrario, las explicaciones reduccionistas constituirían la modalidad explicativa fundamental para cualquier programa que rechace la autonomía ontológica y metodológica de la psicología.

En segundo lugar, parece cuestionable clasificar a tales explicaciones como causales (salvo que se esté utilizando el término 'causal' con una amplitud tal que permita incluir en su extensión a otras clases de relaciones objetivas). La subordinación de fenómenos como la percepción y la inteligencia a estructuras físicas a partir del principio del isomorfismo (como intentara la escuela de la Gestalt), o el intento de explicar los mecanismos de asociación por medio de los

principios del condicionamiento pavloviano, difícilmente parecen constituir explicaciones causales, sino más bien intentos de explicar estructuras de nivel superior en términos de estructuras de nivel inferior; en este sentido, parecería más apropiado recurrir a alguna clase de relación objetiva sincrónica (como algunos autores hicieron al apelar a la idea de emergencia), que vincule ambos niveles de la realidad.

Un último aspecto controvertible de esta concepción de la explicación psicológica es el relativo a los requisitos que las explicaciones deben satisfacer.

Toda explicación, como hemos visto, debe postular una estructura legal de la cual debe deducirse la ley o regularidad empírica a explicar, y tal deducción debe aplicarse a un sustrato real o modelo que se adapta a la deducción y debe representar las relaciones postuladas en ella. Estas parecen ser, entonces, características necesarias de una explicación psicológica 'adecuada'. Si bien hemos visto que Piaget considera que la multiplicidad de clases de explicación en psicología es una situación 'desgraciada', admite la validez de los diferentes tipos expuestos. Sin embargo, aun cuando aceptemos considerarlas necesarias, tales condiciones no parecen ser suficientes. En particular, no establecen requerimientos acerca de las condiciones que las leyes explicativas deben satisfacer. Es visible la falta de la exigencia de las leyes explicativas se hallen bien confirmadas. Como consecuencia, se admiten como explicaciones legítimas las explicaciones psicoanalíticas, particularmente aquellas que apelan a constructos como los de 'relación objetal' y 'energía psíquica' (libido), fuertemente cuestionables.

En síntesis, el admitir distintos tipos de explicación psicológica no nos deja necesariamente en mejor posición con respecto a la necesidad de ajustar ciertas explicaciones debido a la presión ejercida por los argumentos de la exclusión.

A continuación examinaremos un intento de mostrar cómo ciertas explicaciones psicológicas son inmunes a los argumentos de la exclusión.

4. Explicaciones no cartesianas y exclusión causal

Una manera de disminuir el impacto de los argumentos de la exclusión causal sobre la explicación psicológica es mostrar que existen diversas clases de explicación en esta disciplina que no son vulnerables a esa clase de argumentos. Tal es el intento

de Montgomery (1995), quien pretende mostrar cómo, independientemente del éxito que tales argumentos puedan tener sobre ciertos tipos de explicación psicológica, son inocuos contra otros tipos frecuentemente utilizados en psicología cognitiva.

Este autor considera que el modelo dominante de explicación psicológica en el ámbito de la filosofía de la mente es el llamado 'modelo cartesiano', descripto originalmente por el filósofo francés y que sobrevive, con modificaciones menores, en la discusión sobre la causación mental. Pese a este predominio, señala que tal modelo es solamente una parte de lo que puede afirmarse respecto de la explicación psicológica. Existen otros modelos explicativos no cartesianos empleados regularmente en psicología cognitiva que no son afectados por los argumentos contra la relevancia explicativa de las propiedades mentales.

En el modelo cartesiano de explicación, los sucesos psicológicos pueden jugar los siguientes roles: a) como efectos próximos de sucesos no psicológicos en el cerebro y el sistema nervioso; b) como efectos de sucesos en el entorno; c) como causas y como efectos de otros sucesos psicológicos; y d) como causas próximas de sucesos no psicológicos en el sistema nervioso, y a menudo como causas de sucesos más remotos en el resto del cuerpo y en el entorno.

Este modelo cartesiano (o 'psicología cartesiana') contiene además dos condiciones auxiliares que estipulan cuáles sucesos deben ser considerados sucesos psicológicos que pueden jugar los roles especificados en a), b), c) y d). En primer lugar, los sucesos psicológicos son concebidos como la ocurrencia o el cambio en las propiedades de actitudes (proposiciones o no), ideas, representaciones mentales o *qualia* concientes; en segundo lugar, la caracterización admisible de esos sucesos psicológicos es provista por el siguiente principio: toda vez que sea posible, los sucesos psicológicos deben ser caracterizados adhiriendo al principio del solipsismo metodológico, esto es, el principio según el cual ningún estado mental presupone la existencia de otro individuo aparte de aquel al cual se adscribe ese estado. Los apartamientos del principio del solipsismo metodológico son tolerados sólo cuando son necesarios para lograr una caracterización mínimamente completa del contenido del estado mental, caracterización que le permita jugar algunos de los roles establecidos en a)-d).

Estos apartamientos del principio del principio del solipsismo metodológico, que aparentemente se alejan de la doctrina cartesiana, descansan en dos consideraciones. En primer término, señala Montgomery, ya sea que Descartes lo advirtiera o no, si algunas representaciones mentales o actitudes proposicionales son necesariamente no solipsistas, de hecho él mismo las incluía dentro de sus explicaciones psicológicas. En segundo término, el apelar a actitudes proposicionales no solipsistas del tipo que autores como Putnam o Burge han descripto no supone más que un cambio relativamente pequeño del modelo explicativo propio de Descartes. Dos consideraciones sostienen esta afirmación: en primer lugar, todas las versiones del modelo cartesiano descuidan una variedad importante de explicaciones psicológicas utilizadas en la psicología cognitiva contemporánea; en segundo lugar, las explicaciones dentro del modelo cartesiano son vulnerables a los argumentos standard en favor de la ineficacia explicativa de las propiedades psicológicas, mientras que las variedades no cartesianas, como intenta demostrar, no lo son.

Montgomery considera que un rasgo distintivo del modelo cartesiano es que no admite relaciones entre la mente y el mundo, como la verdad y falsedad o la exactitud e inexactitud en el *explanandum* o en el *explanans* de las explicaciones psicológicas; por esta razón, las explicaciones del error cognitivo están fuera del alcance del modelo cartesiano. Sin embargo, los contenidos cognitivos semánticamente evaluados (o 'contenidos fundados' [*grounded content*]) constituyen un rasgo de muchas explicaciones en ciencias cognitivas, tanto en el *explanans* cuanto en el *explanandum*, y aparecen en una variedad de formas explicativas.

En particular, Montgomery describe cuatro variedades de explicaciones cognitivas que apelan al contenido fundado. Estas explicaciones permiten que la exactitud y la inexactitud estén presentes en el *explanandum* y en el *explanans*.

1. La exactitud engendra exactitud. Como ejemplo de esta variedad explicativa Montgomery cita la conocida teorización de David Marr acerca de la visión. En este modelo, señala, se distinguen cuatro etapas sucesivas de representación en el sistema nervioso visual, cada una de las cuales se deriva de la precedente, desde la representación de los valores de intensidad de la luz a lo largo de la superficie de la retina hasta la formación del modelo 3-D, una representación

tridimensional completa de la estructura de los objetos percibidos. El modelo explicativo de Marr no sólo muestra cómo obtenemos información confiablemente precisa acerca del mundo a través de la visión, sino que también está claro que la confiabilidad de la información presente en cada etapa depende de, y es explicada por, la confiabilidad de la información presente en la etapa anterior. Tanto el *explanans* cuanto el *explanandum* apelan al contenido fundado, y exhiben lo que Montgomery denomina 'compromiso explicativo' [*explanatory engagement*]: es la exactitud de la información sobre la profundidad y la orientación lo que es explicado, y es la exactitud del emparejamiento [*matching*] lo que realmente ayuda a explicar.

2. La exactitud engendra inexactitud. Ejemplo de este caso son las explicaciones proporcionadas por Tversky y Kahnemann en sus investigaciones sobre la 'disponibilidad heurística' [*availability heuristic*]. La disponibilidad heurística es una estrategia utilizada para realizar inferencias acerca de frecuencias y probabilidades, que conduce a los sujetos a producir evaluaciones de frecuencias sistemáticamente inexactas sobre la base de información exacta. En los experimentos de Tversky y Kahnemann, la inexactitud en las respuestas de los sujetos al estimar frecuencias es el producto conjunto del recuerdo exacto de cierta información presentada previamente y de un método sistemáticamente sesgado para estimar frecuencias sobre la base de ítemes recordados con exactitud. En opinión de Montgomery, Tversky y Kahnemann no están meramente redescribiendo una secuencia causal cartesiana, sino que están introduciendo *explananda* y *explanans* no cartesianos. Se trata de explicar la inexactitud de las conclusiones de los sujetos sobre la frecuencia, y no sólo la producción de una conclusión acerca de la frecuencia que, como ocurre, resulta ser inexacta.
3. La inexactitud engendra exactitud. Esta posibilidad es ilustrada con los casos de aprendizaje a partir de errores y, en particular, con la simulación conexionista NETtalk diseñada por Sejnowsky y Rosenberg, la cual aprende a pronunciar palabras escritas en inglés. La simulación incorpora una rutina para 1) comparar las representaciones fonéticas incorrectas, y 2) realizar ajustes incrementables en los pesos a través de la red, que tienden a minimizar la extensión del desajuste de

los ensayos subsecuentes. La habilidad de la simulación para producir representaciones fonéticas correctas es explicada conjuntamente por su proclividad a cometer errores y por la habilidad del algoritmo de aprendizaje de retropropagación [*back propagation*] para usar información acerca de la dirección de los errores para un rendimiento correcto en el futuro.

4. La inexactitud engendra inexactitud. Este tipo de fenómeno, y su correspondiente explicación, es ejemplificado por Montgomery con las conocidas investigaciones de Bartlett acerca de los errores en los procesos de evocación de información. En el recuerdo de una narración breve, los sujetos de la experiencia introducen errores en los detalles de la historia. Bartlett argumentó que muchos de los errores de los sujetos fueron el resultado de los intentos inconcientes de tornar la historia más familiar y satisfactoria, introduciendo en ella supuestos derivados del propio trasfondo personal y cultural. En opinión de Montgomery, un examen de la explicación del error proporcionada por Bartlett muestra que el error inicial que conduce a la introducción de estructuras de conocimiento ajenas al relato puede producir diversos tipos de errores subsecuentes.

Montgomery considera que son varias las fuentes a partir de las cuales se originan las preocupaciones relativas a la eficacia causal y la relevancia explicativa de las propiedades mentales; sin embargo, todas ellas han presupuesto el modelo cartesiano de explicación psicológica. Expone dos argumentos que intentan mostrar la carencia de compromiso explicativo que los sucesos psicológicos muestran al tratar de explicar otros sucesos psicológicos. El primero de ellos es el siguiente:

1. Un suceso neurofisiológico es capaz de jugar un rol en la causación de otro suceso neurofisiológico sólo en virtud de las propiedades neurofisiológicas que posee.
2. Los sucesos psicológicos son, como casos, idénticos [*token identical*] a sucesos neurofisiológicos.
3. Las propiedades psicológicas de los sucesos psicológicos no son idénticas, o reductibles a, propiedades neurofisiológicas.

---

4. No es el caso que los sucesos psicológicos sean capaces de jugar un rol en la causación de otros sucesos psicológicos, total o parcialmente, en virtud de las propiedades psicológicas poseídas por los primeros.
5. Las propiedades psicológicas pueden jugar un rol en la explicación de la ocurrencia de sucesos psicológicos sólo si 4. es falsa (*i. e.*, sólo si los sucesos psicológicos *pueden* jugar un rol en la causación de otros sucesos psicológicos, total o parcialmente, en virtud de las propiedades psicológicas poseídas por los primeros).[52]

---

6. Las propiedades psicológicas no pueden jugar un rol en la explicación de la ocurrencia de sucesos psicológicos (p. 229. Cursivas del autor).

El segundo de los argumentos es el siguiente:

1'. Sólo los sucesos neurofisiológicos, sucesos fisiológicos o sucesos físicos son capaces de jugar un rol en la causación de la conducta, y sólo en virtud de las propiedades neurofisiológicas, o fisiológicas u otras propiedades físicas poseídas por esos sucesos.
2'. Las propiedades psicológicas no son idénticas o reductibles a propiedades neurofisiológicas o fisiológicas u otras propiedades físicas del tipo involucrado en la causación de la conducta.

3'. No es el caso que los sucesos psicológicos sean capaces de jugar un rol en la causación de la conducta, total o parcialmente, en virtud de las propiedades psicológicas poseídas por ellos.
4'. Las propiedades psicológicas pueden jugar un rol en la explicación de la ocurrencia de una conducta sólo si 3' es falsa (*i. e.*, sólo si los sucesos psicológicos *pueden* jugar un rol en la causación de la conducta en virtud de las propiedades psicológicas poseídas por esos sucesos psicológicos).

5'. Las propiedades psicológicas no pueden jugar un rol en la explicación de la ocurrencia de conductas (pp. 232-3. Cursivas del autor).

[52] Montgomery señala, con respecto a esta premisa, que a menudo no es explicitada en las discusiones sobre la causación mental, pero que claramente es supuesta de manera tácita. Advierte además que los enfoques no causales de la explicación psicológica, alguna vez predominantes, han sido ampliamente abandonados en favor del punto de vista según el cual si las propiedades psicológicas pueden jugar un rol explicativo, deben hacerlo dentro de explicaciones causales. Es plausible pensar que Montgomery está haciendo referencia al debate 'explicaciones racionalizadoras' *vs.* 'explicaciones causales'; sin embargo, parece aventurado (e injustificado) sostener que la única relevancia explicativa posible para las propiedades psicológicas es la que tiene lugar en explicaciones causales. Por otra parte, parece razonable afirmar que 5) debería restringirse a explicaciones causales; esto es, las propiedades psicológicas sólo serían capaces de jugar un rol en explicaciones *causales* de la ocurrencia de sucesos psicológicos si 4) es falsa. Volveremos luego sobre esta cuestión.

Ambos argumentos, considera Montgomery, son válidos y, *prima facie*, correctos [*sound*] en lo que respecta a las explicaciones que se ajustan al modelo cartesiano. Sin embargo, hay respuestas plausibles para ambos en lo que respecta a las explicaciones que no se ajustan a este modelo.

El primer argumento, señala Montgomery, involucra una estrategia general tendiente a mostrar la inactividad [*inertness*] explicativa de las propiedades mentales. Esta estrategia está destinada a mostrar que el poder explicativo que los sucesos mentales cartesianos parecen tener en virtud de sus propiedades psicológicas dentro del modelo cartesiano, es capturado por el conjunto de propiedades neurofisiológicas que también poseen dichos sucesos. Sin embargo, las explicaciones psicológicas que apelan al contenido fundado no pretenden explicar simplemente sucesos psicológicos estrechos, como ocurren en el modelo cartesiano, sino que en ellas se trata de explicar ciertos sucesos psicológicos híbridos consistentes en la relación semántica entre tres *relata*: un suceso psicológico al estilo cartesiano, la relación semántica que tal estado acarrea, y el mundo real, que determina el *status* semántico de ese contenido. La cuestión es, entonces, si la captura explicativa que parece tener lugar en el caso de los sucesos cartesianos ocurre en el caso de tales sucesos psicológicos híbridos.

Las explicaciones que apelan a contenidos fundados, señala Montgomery, sacan partido de las relaciones semánticas de verdad, falsedad, exactitud e inexactitud que tienen lugar entre los contenidos de los sucesos cartesianos, descriptos en el modo *de dicto*, y el mundo. Ninguna otra propiedad de los individuos, del mundo o de la relación entre ellos parece ser capaz de capturar el rol explicativo jugado por los aspectos relevantes de tales relaciones semánticas. Ninguna referencia a hechos exclusivamente neurofisiológicos servirá para proporcionar una explicación adecuada que sustituya a una explicación que apele a la cognición fundada que sirve para explicar por qué la percepción visual es confiable, o por qué los sujetos de Tversky y Kahneman hacen estimaciones pobres de la frecuencia. Por contraste, observa, en los casos en que los sucesos psicológicos son entendidos al modo cartesiano, las propiedades neurofisiológicas parecen ser capaces de lograr la captura explicativa. Esta posibilidad es la que contribuye a crear el problema acerca de la causación y la explicación mental dentro del modelo

cartesiano. Por el contrario, toda vez que se considera que las explicaciones apelan al contenido fundado, la amenaza de la captura explicativa desaparece. La carga de la prueba recae sobre quienes piensan que existen propiedades capaces de lograr la captura explicativa.

Montgomery considera que también hay respuestas plausibles para el segundo argumento. Podría afirmarse, en primer lugar, que el segundo argumento logra despojar a las explicaciones psicológicas de compromiso explicativo cuando ellas aparecen en las explicaciones de la conducta bajo el modelo cartesiano; sin embargo, ninguno de los tipos de explicaciones descripto intenta ser una explicación de la conducta. Tales explicaciones sólo pretenden explicar la ocurrencia de ciertos sucesos psicológicos híbridos. No obstante, esta argumentación, al desconectar las explicaciones que apelan al contenido fundado de la explicación de la conducta, conlleva el riesgo de condenarlas a la irrelevancia. Es difícil de creer, observa Montgomery, que cualquier clase de explicación psicológica totalmente divorciada de la conducta pueda ser una forma legítima de explicación psicológica.[53] Afortunadamente, prosigue, existen formas explicativas que apelan al contenido fundado que refieren a la conducta.

A diferencia del modelo cartesiano, en el cual se pretende explicar las conductas *simpliciter*, en las explicaciones que apelan al contenido fundado se trata de un tipo diferente de *explanandum*. En particular, se trata de explicar 'el éxito de una persona en lograr su meta o deseo'. En tal caso, el *explanandum* es 'una relación semántica entre una meta o un deseo y alguna conducta (o su efecto ambiental subsecuente) en virtud de la cual la meta o deseo esa meta o deseo es *lograda*' (p. 234. Cursiva del autor). De esta forma, el *explanandum* incluye la conducta, pero la explicación no es el resultado de un intento de explicar la conducta *simpliciter*. El *explanandum* es, visiblemente, un caso de contenido fundado. Podemos explicar la satisfacción de nuestras metas apelando a la exactitud de la información que poseemos, es decir, apelando a otro contenido fundado.[54]

[53] Esta observación es pertinente en relación con aquellas estrategias que se conforman con mantener la causación entre sucesos mentales, descartando la causación de sucesos mentales a sucesos físicos y viceversa. *Cfr.* al respecto el capítulo VI.

[54] Montgomery afirma que su propuesta constituye una variante de la estrategia de los dos *explananda*, ya que 'elimina la competencia explicativa entre explicaciones psicológicas y neurofisiológicas al construirlas como intentos de explicar diferentes cosas' (p. 235). En el capítulo VI examinaremos otra

En opinión de Montgomery, en síntesis, es posible que los argumentos expuestos sean sólidos en lo que respecta a las explicaciones que siguen el modelo cartesiano, pero no lo son con respecto a las explicaciones que apelan al contenido fundado y a propiedades con real compromiso explicativo.

Las réplicas que Montgomery ofrece a ambos argumentos deben analizarse cuidadosamente. Examinemos ahora la respuesta al primer argumento. En primer lugar, y por razones que luego expondremos, parece importante reflexionar sobre la ausencia de una toma de posición con respecto a la cuestión de si las explicaciones de sucesos híbridos que apelan al contenido fundado son o no causales. Montgomery claramente se abstiene de tomar partido con respecto a esta cuestión: 'mi estrategia en este artículo ha sido argumentar por el compromiso explicativo del contenido fundado (...) No he tomado posición, por ejemplo, sobre la cuestión de si esas explicaciones son causales o no causales' (p. 236). No hay razones para exigir que la propuesta se pronuncie de manera concluyente sobre esa cuestión, cuando su objetivo ha sido sólo mostrar como ciertas explicaciones no son vulnerables a determinados argumentos. Sin embargo, es plausible pensar que hay razones más profundas y complejas para el mantenimiento de una posición neutral con respecto a la cuestión.

Hemos visto que, según este argumento, los sucesos psicológicos no juegan ningún rol en la causación de otros sucesos psicológicos en virtud de sus propiedades psicológicas; las propiedades psicológicas sólo pueden jugar un rol en la explicación de sucesos psicológicos en tanto la premisa anterior sea falsa. Ahora bien, según Montgomery el argumento encierra una estrategia general que muestra que los poderes explicativos que los sucesos cartesianos parecen tener en virtud de sus propiedades psicológicas son capturados por las propiedades neurofisiológicas que el suceso posee. Pero notemos que, en el argumento, la captura explicativa es consecuencia de la ineficacia causal: dado que los poderes causales de las

variante reciente de esta estrategia, debida a Ausonio Marras. Ambas propuestas, sin embargo, difieren marcadamente en alcance: mientras la propuesta de Montgomery intenta mostrar como ciertas explicaciones (las explicaciones no cartesianas) no son vulnerables a los argumentos de la exclusión, Marras intenta demostrar cómo los argumentos de la exclusión son inconcluyentes, y cómo la causación mental y la explicación psicológica pueden conservarse plenamente. Por esta razón ambas propuestas son analizadas por separado.

propiedades psicológicas son sustituidos por los poderes causales de las propiedades neurofisiológicas, las primeras carecen de compromiso explicativo. Cuando Montgomery hace referencia a la estrategia general que conduce a la captura explicativa, parece haber un deslizamiento implícito desde la argumentación causal-explicativa a la argumentación puramente explicativa.

Una razón posible para este deslizamiento es que Montgomery desea permanecer neutral con respecto a la cuestión de si las explicaciones psicológicas de sucesos psicológicos híbridos que apelan al contenido fundado son causales. Por una parte, si las explicaciones fueran no causales, el primer argumento no podría constituir una objeción contra ellas; de esta manera, sería trivialmente verdadero que hay explicaciones psicológicas que no se ven afectadas por tales argumentos. Por otra parte, no hay aquí un argumento o principio que apele a la exclusión explicativa, el cual sería necesario para rechazar una de las explicaciones en juego. Una segunda razón, y quizás más decisiva, es que, según la justificación de la premisa 5 del argumento, las propiedades psicológicas sólo pueden jugar un rol en una explicación si esta explicación es causal; la justificación de esta razón que ofrece Montgomery parecería presionar en favor de la posición según la cual las explicaciones son causales.

La situación es diferente con respecto a la respuesta al segundo argumento. No existe aquí el deslizamiento desde el plano causal-explicativo hacia el plano puramente explicativo que hicimos notar respecto de la réplica al primer argumento. Lo que hay es la identificación de un nuevo *explanandum*, recordemos, que consiste en una relación semántica entre una meta o un deseo y cierta conducta (o sus efectos ambientales subsecuentes), por medio de la cual tal meta o deseo es logrado. Esta postulación de un nuevo *explanandum* es suficiente, en opinión de Montgomery, para evitar los efectos del segundo argumento.

Hasta aquí, no parecería haber razones para negar que tanto las explicaciones psicológicas de sucesos psicológicos híbridos cuanto las explicaciones de la relación semántica entre deseos, metas y conductas que apelan al contenido fundado sean causales. Sin embargo, podría haber razones de peso para no admitir que lo sean. Estas razones podrían originarse en uno de los problemas de la causación mental

distinto al problema de la exclusión: el problema de las propiedades mentales extrínsecas.

Kim (1998) ofrece una clara y elegante exposición de este problema que no requiere de supuestos relativos a las teorías computacionales de la mente. Según esta versión del problema, las causas internas de la conducta física de un organismo deben sobrevenir a partir del estado interno total de ese organismo en ese momento. Parece plausible, entonces, suponer que dos organismos que se encuentren en idéntico estado interno en un momento dado emitirán idéntico output motor. Sin embargo (como los célebres argumentos de Putnam y Burge han mostrado) las propiedades semánticas de los estados internos en general no sobrevienen a partir de sus propiedades sincrónicas internas, sino que involucran hechos acerca de las condiciones históricas y ecológicas del organismo. Debido a esto, dos organismos cuyos estados totales en un momento dado posean idénticas propiedades internas podrán diferir con respecto a las propiedades semánticas que ejemplifican, en los contenidos de sus creencias y deseos, en las extensiones de sus predicados homofónicos y en las condiciones de verdad de sus sentencias homofónicas.

Pero esas diferencias semánticas no deberían introducir diferencias en la salida conductual: el duplicado exacto de una persona en la Tierra Gemela creerá que el compuesto XYZ es húmedo, y no que el agua lo es; asimismo, las ranas en la tierra, con estimulación óptica apropiada, tendrán la 'creencia' de que una mosca está cruzando a través de su campo visual, mientras que las ranas de otro planeta en el cual las moscas no existen, idénticamente estimuladas, no tendrán creencias sobre moscas: 'creerán' que un 'schmy' (pequeños murciélagos negros que son su alimento en la tierra gemela) está atravesando su campo visual.

Por lo tanto, que un estado intencional dado de un organismo ejemplifique una cierta propiedad semántica es un hecho relacional, un hecho que involucra de manera esencial las relaciones del organismo con varios factores externos, ambientales e históricos. Pero debido a nuestra creencia de que las propiedades causales implicadas en la producción de la conducta son no relacionales, o intrínsecas, parecería que las únicas propiedades que serían capaces de cumplir esa función son las propiedades sintácticas, dejando a las propiedades semánticas sin un rol causal activo que cumplir. El problema consiste entonces en mostrar cómo las

propiedades extrínsecas, relacionales, pueden ser causalmente eficaces en la producción de la conducta.

Debido a este problema, parecería que hay una muy buena razón para rechazar la afirmación de que las explicaciones basadas en el contenido fundado sean explicaciones causales: si tales explicaciones hacen un uso esencial de ciertas propiedades semánticas presentes en los sucesos híbridos, parecería inevitable que deban enfrentar el problema planteado. Parece difícil negar que las explicaciones deberán postular una cadena causal que vincularía estados psicológicos con contenido entre sí, y entre estados con contenido y la conducta. Las relaciones causales deberían darse entre tales estados entre sí y tales estados y la conducta *en virtud* de sus propiedades de contenido; de esta forma, la explicación del sesgo cognitivo en los sujetos de las investigaciones de Tversky y Kahneman debería darse en términos de cómo un estado psicológico que contiene información exacta *causa* otro estado que contiene información inexacta.

Parece haber presiones, en síntesis, en los dos sentidos: para considerar tales explicaciones como causales y, a la vez, para considerarlas no causales. Si bien podría pensarse que la situación puede ser estable, hay razones para pensar que debe tomarse una posición definida. En particular, ¿en base a qué criterios podríamos evaluar, desde el punto de vista filosófico, la satisfactoriedad de esas explicaciones?[55] Si no nos importara la evaluación filosófica, estaríamos conformes con mostrar que tales explicaciones son utilizadas en psicología cognitiva, pero obviamente no es éste el caso.

Por supuesto, el afirmar que tales explicaciones no son causales no implica aceptar alguna clase de irrealismo explicativo (para emplear la terminología de Kim), que niegue que las explicaciones deban sustentarse en relaciones de determinación

[55] Cummins (1983) señala que una estrategia explicativa tiene una metodología cuando existe un conjunto de condiciones de adecuación para la aplicación de esa estrategia. De esta forma, se tiene un conjunto de principios para distinguir las aplicaciones legítimas de las ilegítimas de tal estrategia. Las preguntas críticas que, en opinión de Cummins, surgen acerca de aplicaciones particulares de una estrategia explicativa son: ¿podría tener la estrategia explicativa tener alguna fuerza explicativa en tal aplicación? y ¿qué tipo de consideraciones evidenciales tenderían a apoyar o socavar una aplicación de la estrategia en este caso? La pregunta que parece pertinente, para el caso de las explicaciones basadas en el contenido fundado de Montgomery, es la siguiente: ¿podemos aplicar ciertos principios de adecuación (los propuestos por Cummins, u otros) sin saber algo más acerca de determinados rasgos sustantivos de tales explicaciones? Parece razonable sospechar que la respuesta es negativa.

objetivas; tales explicaciones podrían fundamentarse en otra clase de relación objetiva que no sea la causalidad. En cualquier caso, existirá una presión para proporcionar una caracterización sustantiva de esas explicaciones.

No carece de interés observar que Montgomery no proporciona razones por las cuales permanecer neutral con respecto al ítem del carácter causal o no causal de las explicaciones que apelan al contenido fundado. Señala, en una de sus observaciones finales, que su propuesta difiere de la de Ruth Millikan en que, a diferencia de ésta, no intenta ofrecer una caracterización sustantiva de las explicaciones que apelan al contenido fundado. Considera que si bien la estrategia de Millikan es más prometedora a largo plazo, en el corto plazo su análisis de las explicaciones sólo genera nuevas controversias. Si las observaciones que hemos expuesto son sólidas, el negarse a tomar posición con respecto a la naturaleza de las explicaciones no logra más que ganar tiempo antes de enfrentar problemas complejos.

Notemos, por último, que los fracasos en mostrar que ciertas explicaciones psicológicas no son explicaciones causales y que no se ven afectadas por los argumentos de la exclusión sólo autorizaría a pensar que *esas* explicaciones pueden verse en problemas; a falta de una caracterización y clasificación sistemática y completa de los diferentes tipos de explicaciones psicológicas, la cuestión relativa a cuales de ellas se ven realmente en dificultades como consecuencia de los argumentos de la exclusión sólo puede quedar, en alguna medida, como una cuestión abierta.

En el siguiente capítulo examinaremos un conjunto de respuestas al problema de la exclusión y la relevancia explicativa, a saber, aquellas propuestas que coinciden en que los sucesos mentales carecen por completo de efectos causales, y que, pese a ello, intentan salvar su rol en las explicaciones de otros sucesos mentales y de la conducta.

CAPÍTULO IV: EPIFENOMENISMO, AISLACIONISMO Y EXPLICACION

1. La alternativa epifenomenista

En el capítulo I se han expuesto sólidos argumentos destinados a probar la ineficacia causal de lo mental, argumentos que parecen conducir directamente hacia el epifenomenismo[56] (y, con la adición de algunas tesis, al irrealismo y al eliminativismo). ¿Es esa posición una alternativa que deba ser considerada seriamente? La respuesta es claramente negativa para la mayoría de los autores. Posiblemente muchos de ellos estarían de acuerdo con Burge (1993) en su afirmación de que no hay reales perspectivas de que el epifenomenismo se constituya en una opción creíble, y que sería mejor considerarlo como un instrumento para clarificar nuestras creencias más profundas.[57] Otros con seguridad suscribirían la frecuentemente citada afirmación de Fodor (quizás sin el tono apocalíptico que éste le impone): 'si no es literalmente cierto que mi deseo es causalmente responsable de mi esfuerzo (...), y mi creencia es causalmente responsable de mi dichos, (...) si nada de esto es literalmente verdadero, entonces prácticamente todo lo que creo acerca de cualquier cosa es falso, y es el fin del mundo' (Fodor, 1989, p. 156).

Lo que estas posiciones plantean es, de alguna forma, la cuestión relativa a si el epifenomenismo puede ser parte de la solución al problema de la causación mental o si, por el contrario, es parte del propio problema.

Si, de alguna manera, los argumentos en favor del epifenomenismo resultaran concluyentes, no habría más alternativa que aceptarlo, aun con todas las consecuencias que esto implicaría (en primer lugar, la aparente pérdida de las explicaciones causales basadas en propiedades mentales). Tal situación de pérdida de plausibilidad de nuestros conceptos explicativos de sentido común no resultaría

---

[56] Conviene observar que, en rigor, algunas de las posiciones que suelen ser calificadas de epifenomenistas no lo serían si nos atenemos a la definición clásica de esta doctrina. Para la mayoría de estas posiciones lo mental no es causado por lo físico, sino que sobreviene o es realizado por él; sin embargo, lo mental carecería de poderes causales, por lo cual la semejanza con la definición clásica del epifenomenismo es suficiente como para que pueda considerárselas conjuntamente con posiciones epifenomenistas: lo mental es determinado por lo físico, pero en sí mismo carece del poder para producir nada.

desconocida. Por ejemplo, la relatividad y la mecánica cuántica han trastornado de manera radical nuestras maneras de concebir el mundo físico que nos rodea; aun cuando no comprendamos muy bien muchas de sus implicaciones (especialmente en el caso de la cuántica), estamos obligados a admitir que estas teorías, y no nuestro sentido común, proporcionan una imagen ajustada del mundo físico. El grado de confirmación que esas teorías poseen nos obliga a aceptarlas, aun al costo de renunciar a la creencia en la corrección de nuestra imagen de sentido común del mundo físico. Pero la situación con respecto al epifenomenismo no es la misma. Muchos de los argumentos existentes no prueban que el epifenomenismo sea la alternativa obligada; sólo apoyan la conclusión de que el marco materialista no reduccionista parece conducir a esta doctrina. Sin embargo, también puede conducir al dualismo fuerte o al eliminativismo.[58] Al analizar la alternativa epifenomenista también podemos evaluar sus consecuencias. Si el epifenomenismo conlleva, como parece inexorablemente a hacerlo, el abandono de nuestras explicaciones causales (tanto de sentido común cuanto científicas) basadas en estados y procesos mentales, podría considerarse que esta es una razón suficientemente buena para concluir que tal doctrina es inaceptable y debe ser rechazada. Sin embargo, se plantea aquí la conocida situación en la cual lo que es un *modus ponens* para un filósofo es un *modus tollens* para otro: si para algunos el epifenomenismo debe ser rechazado porque implica el abandono de buena parte de nuestro acervo de explicaciones psicológicas, para otros su aceptación conduce necesariamente a tal consecuencia.

Hay autores que, aun a su pesar, se ven llevados a aceptar la ineficacia causal de lo mental, debido a la fuerza de los argumentos que conducen en esta dirección; esta aceptación se ve en ocasiones acompañada del intento de mitigar algunas de sus consecuencias más indeseables. En este capítulo examinaremos principalmente una de las cuestiones que suscita la aceptación de la ineficacia causal de la mente, esto es, las consecuencias que, en el plano explicativo, tendría esta doctrina. Este análisis implicará partir del supuesto de que nos veríamos conducidos por necesidad a la doctrina de la ineficacia causal de la mente, y respecto de la cual sólo se puede intentar atenuar algunas de sus consecuencias más indeseables. Esto requerirá, en

[57] Tal vez muchos dirían respecto de esta doctrina lo mismo que Mary Hesse ha observado respecto de la concepción instrumentalista de las teorías científicas: 'es más comentada que creída'.

primer lugar, distinguir entre diferentes tipos de epifenomenismo. Luego examinaremos la posibilidad de que el epifenomenismo sea vulnerable a objeciones basadas en el éxito explicativo de la psicología. Esta será una primera forma en la que el epifenomenismo se vincula con las cuestiones epistemológicas vinculadas a las explicaciones psicológicas. Por último, se analizarán tres posiciones que admiten, como mínimo, que lo mental no tiene efectos causales en el mundo físico, y que, pese a ello, coinciden en la importancia o necesidad de mantener lo mental como base para la mejora de nuestras explicaciones o nuestra comprensión del mundo. Estas dos formas en que el epifenomenismo se vincula con las cuestiones ontológicas y gnoseológicas suscitadas por las explicaciones psicológicas no se encuentran, como veremos, desconectadas entre sí.

Es importante observar que la articulación de un argumento sólido que permita preservar las explicaciones psicológicas, en particular las explicaciones científicas en psicología, es de gran importancia teórica, ya que contribuiría a anular una de las dos consecuencias principales del argumento de la exclusión: la irrelevancia explicativa.

Sin embargo, no todas las propuestas que expondremos tienen los mismos objetivos: mientras que algunas sólo pretenden mostrar cómo podemos seguir utilizando las explicaciones que hacen referencia a estados mentales, sin intentar mostrar cómo las explicaciones psicológicas pueden funcionar en ausencia de poderes causales de las entidades que postulan, otras intentan salvar las explicaciones psicológicas de lo mental (junto, en algunos casos, con las de las ciencias especiales) en el contexto científico. Debido a esto, no enfrentan las mismas objeciones. Respecto de estas últimas propondremos una restricción general: la manera en que se interprete a las explicaciones de la conducta y de los procesos mentales en el marco de una ontología epifenomenista debe preservar algunas de las características y poderes centrales que normalmente atribuimos a las explicaciones científicas en psicología. Concluiremos que presentan ciertas limitaciones que parecen impedirles lograr este objetivo.

[58] *Cfr.* Kim (1998), pp. 119-120.

2. Dos tipos básicos de epifenomenismo

Conviene comenzar observando que el epifenomenismo no es una doctrina cuyo alcance se limite al ámbito de los hechos mentales. En general, las propiedades de todas las ciencias especiales pueden ver amenazada su eficacia causal. Sin embargo, donde se ha planteado de manera más acuciante la amenaza del epifenomenismo es en el ámbito de los fenómenos mentales.

La doctrina clásica del epifenomenismo referente a lo psíquico sostiene que los fenómenos mentales no tienen efectos causales: son causalmente impotentes o inertes.[59] Sólo los fenómenos físicos pueden ser causalmente eficaces. Los fenómenos mentales son efectos de fenómenos físicos, los cuales constituyen sus únicas causas. Sin embargo, al discutir el epifenomenismo en el marco del fisicalismo contemporáneo, se hace necesario precisar y matizar la definición clásica.

De acuerdo con McLaughlin (1989, 1994), es necesario distinguir dos clases de epifenomenismo: epifenomenismo de casos [*token epiphenomenalism*] y epifenomenismo de tipos [*type epiphenomenalism*].[60] El epifenomenismo de casos sostiene que los sucesos mentales no pueden causar nada, mientras que el epifenomenismo de tipos sostiene que ningún suceso puede causar nada en virtud de caer bajo un tipo mental. En otros términos:

1. *Epifenomenismo de tipos*: a) Los sucesos pueden ser causas en virtud de caer bajo tipos físicos, pero b) los sucesos no pueden ser causas en virtud de caer bajo tipos mentales.
2. *Epifenomenismo de casos*: a) los sucesos físicos pueden causar sucesos mentales, pero b) los sucesos mentales no pueden causar nada.

[59] Según Dennet (1991), el término 'epifenomenismo' es empleado con significados completamente distintos en psicología, ciencias cognitivas y filosofía. Atribuye a Huxley la definición del término de uso más común en psicología, disciplina en la cual significa 'propiedad no funcional' o 'producto derivado' (por ejemplo, el hecho de que las personas se muerdan los labios o den golpecitos con los pies mientras piensan). Tales productos derivados no juegan ningún papel operativo ni estructural en los procesos de pensamiento: son no funcionales. Pero los epifenómenos, en este sentido, si bien carecen de importancia funcional, tienen numerosos efectos en el mundo (entre otros, dar golpecitos produce un sonido que se puede grabar). El significado filosófico, por el contrario, hace referencia a fenómenos que son efectos, pero que por sí mismos carecen de cualquier tipo de efecto sobre el mundo físico.

[60] Distinción inspirada, según McLaughlin, en Broad.

Esta distinción surge en principio de una temprana réplica a uno de los principales argumentos epifenomenistas, esto es, el de la ausencia de 'huecos' [*gaps*] en los mecanismos causales neurofisiológicos, ausencia que tornaría causalmente impotentes a los estados mentales. Tal réplica se basó en la idea de que los sucesos mentales no son cambios en (y estados de) una sustancia inmaterial cartesiana, sino cambios en (y estados de) el cerebro. De acuerdo con este enfoque, un suceso determinado puede ser tanto un caso de un tipo neurofisiológico cuanto un caso de un tipo mental, siendo entonces a la vez un suceso físico *y* un suceso mental.

Sin embargo, advierte McLaughlin, esta réplica conduce a plantear la cuestión de la relevancia de las propiedades o tipos mentales en las relaciones causales. Esto es, surge la pregunta acerca de si los sucesos mentales son causas *en virtud* de sus propiedades mentales. El epifenomenismo de tipos es, por lo tanto, una respuesta negativa a este interrogante.

El epifenomenismo de casos, sostiene McLaughlin, implica al epifenomenismo de tipos. Si un suceso puede causar otro en virtud de caer bajo un tipo mental, entonces un suceso podría ser tanto un suceso mental como una causa, con lo que el epifenomenismo de casos resultaría falso. De esto se sigue que si los sucesos mentales no pueden ser causas, entonces los sucesos no pueden ser causas en virtud de caer bajo tipos mentales. No obstante, el rechazo del epifenomenismo de casos no implica el rechazo del epifenomenismo de tipos, si es el caso que un suceso mental pueda ser un suceso físico que posea efectos causales. Si esto fuese así, entonces el epifenomenismo de casos sería falso, pero el epifenomenismo de tipos podría ser verdadero, ya que los sucesos mentales podrían ser causas en virtud de caer bajo tipos físicos, pero no en virtud de caer bajo tipos mentales. Esta relación obliga a examinar de manera independiente los argumentos conducentes a ambas clases de epifenomenismo.

Parece relativamente obvio que ambos tipos de epifenomenismo no enfrentan las mismas clases de objeciones.[61] A continuación examinaremos una de estas objeciones, basada en el éxito explicativo de la psicología.

[61] Algo similar ocurre con respecto a otras clasificaciones del epifenomenismo. Horowitz (1999) distingue el 'epifenomenismo clásico' (la tesis de que los sucesos mentales carecen absolutamente de poderes causales) del 'fisicalismo epifenomenista' (la conjunción de la tesis del fisicalismo de tipos y de la tesis de que los sucesos mentales ejercitan poderes causales en virtud de sus propiedades mentales), y

3. El epifenomenismo y el éxito explicativo de la psicología

Hemos observado al inicio de este capítulo que existen (al menos) dos maneras en que el análisis del epifenomenismo se conecta con las cuestiones epistemológicas relativas a las explicaciones psicológicas: en primer lugar, la manera en que la aceptación del epifenomenismo afecta tanto a nuestras explicaciones psicológicas de sentido común cuanto a las explicaciones de la psicología científica; en segundo lugar, la posibilidad de que el éxito de las explicaciones psicológicas constituya un argumento en contra de la aceptabilidad del epifenomenismo.

Horowitz (1999), ha observado que uno de los argumentos en contra del epifenomenismo clásico ha sido la llamada 'objeción explicativa'. Como hemos expuesto, la objeción señala aproximadamente lo siguiente: el éxito de la psicología en explicar y predecir la conducta implica (o al menos es la mejor explicación) de que las propiedades a las que refiere esa disciplina son causalmente eficaces. Horowitz advierte que esta objeción no es eficaz contra el epifenomenismo de tipos. El éxito predictivo y explicativo no puede ser considerado una razón para considerar esas propiedades causalmente eficaces. Es completamente distinto, observa, considerar que el éxito predictivo y explicativo implica que las entidades o sucesos a los que hace referencia la práctica implica que tales entidades o sucesos existen, que concluir a partir de allí que son causalmente eficaces.

La objeción explicativa, observa Horowitz, presupone una concepción realista, de acuerdo con la cual el éxito predictivo y explicativo de una práctica que hace referencia a propiedades de cierta clase implica que esas propiedades son

---

señala que las objeciones que se han elevado contra el primero no son pertinentes contra el segundo. Estas objeciones contra el epifenomenismo tradicional son cinco: la 'objeción de la introspección' (esto es, el epifenomenismo clásico parece en conflicto con nuestro conocimiento 'desde adentro' de que nuestra mentalidad hace una diferencia); la 'objeción ontológica' (esto es, la posesión de poderes causales es una condición necesaria de existencia, por lo cual, si lo mental existe, debe poseer poderes causales); la 'objeción explicativa' (esto es, que ciertos tipos de actividad humana requieren la existencia de causas no físicas –cuestión ya observada por Descartes-); la 'objeción de las otras mentes' (esto es, el epifenomenismo clásico parece tener dificultades para enfrentar el problema de las otras mentes –para esta posición no está disponible el argumento de que podemos inferir la existencia de tales estados a partir de la conducta de los demás, ya que parece presuponer que tal conducta está causada por sus estados mentales, suposición vedada para el epifenomenista-); la 'objeción evolucionista' (la más reciente de todas, por la cual se sostiene que parece improbable que un organismo tenga cierto rasgo a menos que tal rasgo dote al organismo de alguna ventaja evolutiva). Volveremos a continuación sobre la objeción explicativa. Observemos, de paso, que la fuerza de la objeción evolucionista es, en el mejor de los casos, sumamente controvertida. Distintos autores han señalado que muchos organismos poseen rasgos que,

poseídas por las entidades o sucesos y que son causalmente eficaces. Si bien acuerda con que el éxito reclama algún tipo de explicación realista, señala que podemos estar satisfechos con una explicación que relacione sistemáticamente esas propiedades –propiedades mentales en este caso- con otras propiedades que, se supone, son causalmente eficaces –propiedades del cerebro en este caso-. Esta relación no necesita ser una relación de superveniencia. La existencia de una correlación entre ejemplificaciones de propiedades del cerebro y ejemplificaciones de propiedades mentales es suficiente para explicar como es posible inferir la conducta sobre la base de la posesión de propiedades del último tipo.

Sin embargo, observa Horowitz, esta explicación puede resultar menos aceptable según se acepte uno u otro enfoque acerca de la naturaleza de una explicación correcta. Si se adopta el enfoque de acuerdo con el cual la explicación es un análogo *post facto* de la predicción,[62] lo mental *qua* mental posee poder explicativo en el marco del fisicalismo epifenomenista, dado que podemos predecir la conducta sobre la base de la adscripción de propiedades mentales. Pero si se adoptan enfoques alternativos, la situación es diferente. En el caso de las explicaciones intencionales, los sucesos relacionados no sólo están relacionados causalmente, sino también intencionalmente. Cuando un suceso es causado por otro, el efecto y la causa están semánticamente relacionados el uno con el otro. Las relaciones semánticas tienen lugar entre causas mentales y efectos mentales (la creencia de que *p* y la creencia de que *si p entonces q* causa, y está semánticamente relacionada, con la creencia de que *q*) y entre causas mentales y efectos conductuales. Partiendo del supuesto de que tales explicaciones son exitosas, señala Horowitz, ¿no debería admitirse que tales propiedades intencionales son responsables del rol causal jugado por los estados mentales, y que entonces lo mental *qua* mental es causalmente eficaz?

Horowitz considera que la respuesta a este interrogante es negativa. En su opinión, el poder explicativo de lo mental *qua* mental es independiente de si las propiedades mentales son causalmente eficaces. El éxito de las explicaciones que

---

desde el punto de vista adaptativo, son 'neutros', esto es, no proporcionan ventajas a su poseedor en su ajuste al medio.

[62] Aunque Horowitz no lo menciona, es de suponer que se está refiriendo a los modelos de cobertura legal.

relacionan causalmente sucesos en términos de sus propiedades intencionales no presupone que las relaciones causales tengan lugar en virtud de esas propiedades. Todo lo que presupone es que hay una correlación sistemática entre propiedades causales de los sucesos y sus propiedades intencionales. Y a menos que actuemos *en virtud de* las propiedades intencionales (o, en otros términos, *movidos* por esas razones), no actuamos realmente por razones y no somos racionales y moralmente responsables.

Sólo haremos aquí un breve comentario, ya que la posibilidad de que el epifenomenismo resulte afectado por esta clase de argumentos no es nuestro interés principal en este capítulo.

Debe señalarse que la noción de 'éxito explicativo' es bastante problemática. Podría afirmarse que la psicología, considerada globalmente, tiene éxito explicativo, si se considera la cantidad y variedad de hechos explicados dentro de su marco, hechos que carecen de una explicación dentro de otras disciplinas y dentro de la psicología de sentido común. Veremos en el capítulo V que esta manera de concebir el éxito explicativo tiene consecuencias indeseables. Por ahora, digamos solamente que parece innegable que muchas explicaciones psicológicas presuponen la eficacia causal de las propiedades mentales. Si bien no hay acuerdo interno en la disciplina acerca de cuáles son las características de una explicación correcta, y las diversas corrientes teóricas proponen diferentes criterios de adecuación de las explicaciones,[63] parece un rasgo general de muchos tipos de explicaciones el que los estados mentales poseen eficacia causal.[64] La existencia de explicaciones en competencia, tan frecuente en el ámbito de la psicología, no es un obstáculo para suponer como condición (necesaria, pero seguramente no suficiente) del éxito explicativo la eficacia causal de lo mental. El éxito explicativo, entonces, consistiría en explicar las conductas y los estados psíquicos considerándolos como efectos (al menos, parcialmente) de otros estados psíquicos. Si el epifenomenismo de tipos fuese verdadero, entonces la psicología carecería, en gran medida, de éxito

[63] Con respecto a la existencia de múltiples modelos explicativos en el ámbito de la psicología, véase la descripción proporcionada por Piaget (*supra*, pp. 71-77). La descripción de Piaget debería ser enriquecida con una caracterización de los modelos explicativos cognitivistas desarrollados en las últimas décadas.
[64] Véanse al respecto las consideraciones sobre la teoría de la acción razonada expuestas en el capítulo III.

explicativo, ya que sus explicaciones se basarían en presupuestos erróneos (esto es, en la eficacia causal de los tipos mentales). *Mutatis mutandis*, la situación se asemejaría a la planteada en los ámbitos de la física y la química con la existencia del éter y el flogisto respectivamente: ninguna explicación basada en propiedades de esas entidades podría resultar satisfactoria, ya que tales entidades no existen ni han existido.[65] La posición de Horowitz es que el éxito explicativo de la psicología puede ser explicado de una manera que no sea incompatible con el epifenomenismo de tipos. Por el contrario, nos parece que hay razones para sostener que el epifenomenismo de tipos torna inexistente una parte considerable del éxito explicativo de esa disciplina.

Si el epifenomenismo fuese verdadero, por lo tanto, correspondería adoptar una posición 'reformativa' con respecto a las explicaciones psicológicas. Esto es, habría que explicar la manera en la cual las explicaciones psicológicas, convenientemente reinterpretadas, podrían subsistir en un marco ontológico en el cual lo mental carecería de poderes causales. En los apartados que siguen examinaremos algunas propuestas que exploran este camino.

4. Bieri: la aceptación plena del epifenomenismo

Bieri (1992) considera al epifenomenismo como una opción de aceptación casi obligada, habida cuenta de la solidez de los argumentos que muestran la ineficacia causal de la mente. Este autor comienza exponiendo las implicaciones del epifenomenismo por las cuales no parece realmente una alternativa plausible; aunque son bien conocidas, conviene describirlas a fin de clarificar plenamente lo que la aceptación de esta doctrina involucra (y a fin de poder evaluar cabalmente los argumentos que se ofrezcan en favor de la conservación de las explicaciones psicológicas dentro de un marco epifenomenista):

> Consideraré al epifenomenismo como la doctrina de que nuestra mente, si bien perfectamente real, no determina causalmente nuestra conducta. Las implicaciones de esta doctrina parecen inaceptables. Primero, significaría que vivimos con un error permanente, masivo e irresistible acerca de nosotros mismos. Segundo, aceptar la irrelevancia causal de la mente equivaldría, parece, a una forma de auto-alienación: nuestros cuerpos hacen lo que hacen independientemente de lo que pensemos y

[65] Se hace aquí pertinente la crítica a la posición de Laudan. *Infra*, p. 144.

> sintamos y deseemos. Tercero, si los fenómenos mentales no son antecedentes causales de la conducta, la referencia a ellos no puede constituir una explicación causal de la conducta. Dado que, de hecho, no hay otras explicaciones causales disponibles actualmente, al menos en muchos casos la verdad del epifenomenismo significaría que carecemos en muy gran medida de comprensión de las fuerzas en movimiento detrás de nuestra conducta y la de otros. Cuarto, se dice que el epifenomenismo hace lucir como un milagro a la predictibilidad real de nuestra conducta. Si nuestra estrategia predictiva ordinaria de referirnos al perfil mental de una persona fracasara en identificar los factores causalmente relevantes, el éxito de esta estrategia parecería una chiripa permanentemente repetida.[66] En vista de tales implicaciones ha devenido casi una constricción general de una teoría aceptable de la mente el que sea capaz de evitar el epifenomenismo (p. 283).

La mayor parte del artículo está destinada a exponer las distintas vías por las cuales se llega al epifenomenismo y a la crítica de ciertos argumentos destinados a cerrar el camino a su aceptación.[67] En términos generales, Bieri considera que los argumentos que conducen al epifenomenismo son mucho más sólidos que los que tratan de evitar este resultado. Bieri destina la segunda parte del trabajo, mucho más breve, a discutir ciertas consecuencias que el epifenomenismo tendría para la comprensión psicológica. Estas observaciones son 'breves y algo impresionistas', en palabras del propio autor; sin embargo, manifiesta la esperanza de haber podido mostrar que la situación no sería tan mala si no existiese la causación mental. No hay necesidad, sostiene, de evitar el epifenomenismo a toda costa.

En el resto de este apartado expondremos, en primer lugar, las consideraciones de Bieri referentes a las formas de interpretar las explicaciones psicológicas en el marco del epifenomenismo y, en segundo lugar, comentaremos los que, a nuestro criterio, constituyen aspectos cuestionables de su posición. Dado que, según él mismo admite, sus observaciones son más bien breves y esquemáticas, nos limitaremos a señalar los aspectos que parecen requerir de un análisis mayor para resultar aceptables.

---

[66] Esta afirmación resulta algo llamativa, dada la posibilidad de predecir la conducta mediante enunciados de correlación. Esta estrategia, bien conocida en el ámbito de la ciencia y, por supuesto, también en el de la psicología científica, permite la predicción de la conducta en base a enunciados generales que establecen meras covariaciones entre propiedades o atributos. En el caso de que el epifenomenismo fuese verdadero, podríamos considerar a muchas de nuestras actuales generalizaciones que ligan causalmente estados mentales y conducta como enunciados de correlación, los cuales mantendrían su poder predictivo, si bien perderían su poder explicativo.

[67] En particular, las expuestas por Dretske, Fodor, Kim (su propuesta de la causación superveniente) y Loewer y LePore.

4.1. Tres formas de comprensión psicológica

Responder a las preguntas relativas al rol que los sucesos mentales jugarían en la comprensión psicológica en caso de que el epifenomenismo fuese correcto requiere, según Bieri, distinguir tres formas de comprensión psicológica:

a. La comprensión de la conducta a través explicaciones causales en términos de sucesos mentales.
b. Comprender la conducta a través de interpretaciones racionalizadoras, y
c. Comprender a las personas y sus perfiles psicológicos a través de 'proyección y nueva representación' [*reenactment*].

Estas maneras diferentes de comprensión están definidas por distintos déficits epistémicos y por diferentes expectativas acerca de lo que pueden proporcionar.

a. La verdad del epifenomenismo no implica, para Bieri, que las explicaciones psicológicas no puedan ser vistas como haciendo referencia a causas internas de la conducta. Pero sí implica que estas causas no se encuentran dentro de las propiedades causalmente relevantes para la producción de esa conducta.[68] Como consecuencia, observa, las explicaciones psicológicas no son 'extensibles'.[69] La irrelevancia causal de los contenidos mentales implica que los condicionales contrafácticos mentales son falsos: la conducta no depende contrafácticamente de contenidos mentales.[70] La comprensión causal de la conducta no puede tener lugar en términos de contenidos mentales, sino que debe ser desarrollada en términos fisiológicos.

[68] Bieri considera a los sucesos como 'ejemplificaciones de racimos de propiedades que se extienden espacio-temporalmente' (p. 284).
[69] 'Cuando una explicación de un suceso que comienza citando una regularidad rústica y no básica procede a especificar más y más los mecanismos finos involucrados en la causación, podemos decir que la explicación original está siendo *extendida*, y denominaré a una explicación que, en principio, sea capaz de ser extendida, una explicación *extensible*' (p. 293).
[70] Esta afirmación resulta discutible. Los contrafácticos mentales podrían ser verdaderos en caso de que los estados mentales a los que hacen referencia estuvieran conectados nómicamente con estados fisiológicos eficaces en la producción de la conducta. Si un tipo de proceso mental *M* acompañara invariablemente a un tipo de proceso fisiológico *F*, causalmente eficaz en la producción de una conducta determinada, entonces sería correcto afirmar que, en ausencia del proceso mental *M* la conducta no se habría producido ya que, en ausencia de *M*, tampoco habría ocurrido *F*. (*Cfr.* al respecto la noción de 'equivalentes nómicos simultáneos' propuesta por Goldman. *Supra*, p. 19).

Esta conclusión, prosigue Bieri, puede parecer chocante, dada la escasez de conocimientos fisiológicos relevantes. Sin embargo, considera que la resistencia intuitiva al epifenomenismo no es en realidad tan completa como a menudo se afirma. Bieri ofrece dos razones en apoyo de esta afirmación. En primer lugar, estamos más que dispuestos a decir cosas como 'el whisky lo hizo decir eso' o 'la droga causó que hiciera eso'. Tales casos comparten con los casos normales dos aspectos cruciales: hay un contenido mental y hay una cadena causal fisiológica completa.

Por otra parte, señala que es un hecho notable el que en contextos comunes no nos perturbe el hecho de que las explicaciones mentales no son extensibles, mientras una deficiencia similar nos molesta en el caso de explicaciones causales no mentales. Bieri sugiere que esto se debe a que el rol causal del contenido no es lo que nos interesa primariamente respecto de lo mental, aun cuando no estemos preocupados por la verdad del epifenomenismo; lo que es realmente importante es el rol que el contenido mental juega en los otros dos modos de comprensión psicológica.

b. 'Entender la conducta de alguien' puede significar darle sentido representándola como un fragmento de conducta racional. 'Racionalizar' la conducta significaría atribuir ciertas razones al agente. Pero si el epifenomenismo es verdadero, los contenidos de esas razones resultarían irrelevantes para una descripción causal de la conducta. Si esto es así, contrafácticos como 'si X no hubiese pensado, sentido, deseado o recordado ciertas cosas, no habría ocurrido el hecho Y' serían falsos. Sin embargo, sostiene Bieri, la falsedad de contrafácticos como el anterior *no implica* la falsedad de contrafácticos como 'Si X no hubiera pensado, sentido, deseado y recordado ciertas cosas, no habría sido *racional* para él hacer Y'.

Los condicionales del primer tipo, en tanto condicionales causales, especifican las condiciones para la ocurrencia de un suceso, mientras que los del segundo tipo especifican las condiciones para que éste sea racional. Mientras que el epifenomenismo implica la falsedad de los del primer tipo, no implica la falsedad de los del segundo. Sólo habría un conflicto en el caso en el que se considerara que la conducta, para ser racional, debe ser causada por el contenido mental de las razones

que mueven a la acción. Bieri no considera que haya razones para tal supuesto; por el contrario, observa, las explicaciones racionalizadoras pueden tener éxito independientemente de cualquier pregunta relativa a mecanismos internos. No son las propiedades causales de las razones las que son invocadas en tales explicaciones sino las propiedades de contenido pertinentes para la posibilidad de justificar la conducta. Entonces, concluye, carece de importancia el hecho de que las explicaciones racionalizadoras no sean extensibles: no se supone que lo sean. Por lo tanto, el epifenomenismo no pone en peligro la comprensión psicológica como 'hermenéutica racionalista'.

c. Las personas son, en una medida considerable, irracionales. Sus creencias no se conectan de manera adecuada, sus deseos entran en conflicto y su conducta está guiada por ambivalencias emocionales complejas. Representarlas como agentes racionales, si bien crea una impresión superficial de comprensión, es frecuentemente una distorsión.

Si el epifenomenismo fuese verdadero, sostiene Bieri, no podríamos dar una descripción causal de la irracionalidad en términos de contenido mental. Sin embargo, hay algo que sí podría hacerse: proyectarnos a nosotros mismos dentro de las situaciones de otras personas. Esto significa 'reconstruir sus deliberaciones y por consiguiente entender sus conductas como racionales' (p. 306). Pero también puede hacerse algo muy diferente: 'representar [*reenacting*] o simular en uno mismo el perfil mental de otras personas, incluyendo sus rasgos irracionales' (*ibid.*). Si, al poner en práctica esta posibilidad, llegamos a darnos cuenta de que en tal situación podríamos sentir, pensar y actuar del mismo modo irracional, el déficit de comprensión desaparece.

La representación no sólo es crucial en la comprensión de la irracionalidad, sostiene Bieri, sino que también puede ser decisiva en la comprensión de conductas perfectamente racionales, pero construidas sobre deseos, motivos y emociones inusuales. Las personas a menudo inventan las metas que desean perseguir, metas que pueden ser muy idiosincrásicas, por lo que la nueva representación es crucial para la comprensión.

Este tercer modo de comprensión psicológica, sostiene Bieri, no está en conflicto con el epifenomenismo. Lo que importa es que nos experimentamos a nosotros mismos como mentalmente similares a las personas que son comprendidas. Carece de interés el que las experiencias alcanzadas a través de la representación causen la conducta.

Así, concluye, el segundo y tercer modo de comprensión psicológica son igualmente neutrales con respecto al problema de la causación mental, de manera que no perderían nada de su significación en caso de que el epifenomenismo fuese verdadero.

4.2. La reformulación de las explicaciones psicológicas: una crítica

La posición de Bieri podría ser considerada 'reformativa' con respecto a las explicaciones psicológicas. Esto es, se trataría de mostrar que las explicaciones psicológicas de diversas clases tienen un valor cognoscitivo aun en ausencia de causación mental. No obstante, los argumentos en favor de esta posibilidad nos parecen menos que satisfactorios.

Observemos en primer lugar que Bieri no hace referencia explícitamente a las explicaciones científicas en psicología que apelen a la eficacia causal de los sucesos o propiedades mentales. Dado que no hay un intento explícito de salvar tales explicaciones, consideraremos que está haciendo referencia a las explicaciones psicológicas de sentido común.

En primer lugar, conviene discutir cuáles son los 'contextos comunes' en los cuales la no extensibilidad de las explicaciones mentales no nos perturba. Entenderemos por tal (interpretación que parece la más razonable) los contextos de la vida cotidiana, el dominio natural de la psicología de sentido común. Pero, en este contexto, la pretensión de Bieri parece al menos cuestionable. En el ámbito de la psicología de sentido común, a juzgar tanto por las opiniones de otros autores,[71] cuanto por los estudios en el campo de la psicología social acerca de las

[71] 'Los tipos mentales figuran en nuestras *explicaciones causales cotidianas* de la conducta, de la acción intencional, de la memoria y del razonamiento' (McLaughlin, 1994, p. 281. Cursivas nuestras). '[L]a psicología ordinaria postula estados mentales provistos de contenido como causas de la conducta (Engel, 1988, p. 13).

explicaciones causales de sentido común,[72] resulta más que dudoso que lo mental no juegue un rol crítico en la estructura causal-explicativa de las acciones humanas. Si bien la existencia de opiniones divergentes sobre esta cuestión no prueba que la posición sostenida por Bieri sea necesariamente errónea, resulta suficiente para rechazar como injustificada su pretensión de que las explicaciones en 'contextos ordinarios' no apelan a estados mentales como causas internas de la conducta. Por otra parte, y aunque esta consideración sea tangencial al punto en discusión, tampoco es aceptable sin más que la no extensibilidad, entendida como una deficiencia, nos perturbe en el caso de explicaciones de sucesos que no involucren aspectos mentales. Los sociólogos del conocimiento han enfatizado ciertas características del conocimiento de sentido común que parecen desmentir la presunta insatisfactoriedad de las explicaciones no extensibles. En particular, el conocimiento de sentido común es considerado un *conocimiento práctico*, que orienta la actividad humana en el mundo de la vida cotidiana, y que no es cuestionado a menos que se presente como problemático.[73] Al igual que en el caso anterior, la existencia de opiniones discrepantes acerca de esta cuestión no prueba que la posición de Bieri sea incorrecta, pero limita severamente su plausibilidad sin añadir argumentos más fuertes en su favor.

Tampoco parece cierto que el hecho de que estemos perfectamente dispuestos a aceptar enunciados como 'el whisky causó que dijera eso' sea prueba de que estemos dispuestos a aceptar el epifenomenismo. Por una parte, la aceptación de enunciados de esa clase es compatible con la aceptación de un dualismo fuerte: podemos estar igualmente persuadidos de que 'el deseo de *beber whisky*' causa que

[72] *Cfr*. al respecto Hewstone (1989).

[73] 'Como la vida cotidiana está dominada por el motivo pragmático [la atención que se presta al mundo está determinada por lo que se hace en él], el conocimiento de receta, o sea, el conocimiento que se limita a la competencia pragmática en quehaceres rutinarios ocupa un lugar prominente en el cúmulo social de conocimiento. Por ejemplo, uso el teléfono todos los días con propósitos específicamente pragmáticos de mi incumbencia. Sé cómo hacerlo. También sé que hay que hacer si mi teléfono funciona mal, lo que no significa que sepa cómo repararlo, pero sí que sé a quién hay que recurrir en ese caso (...) No me interesa *por qué* el teléfono funciona de esa manera, ni la enorme cantidad de conocimientos científicos y técnicos que posibilitan la construcción de teléfonos (...) *Mutatis mutandis*, gran parte del cúmulo social de conocimiento consiste en recetas para resolver problemas de rutina. En particular, me interesa poco traspasar el límite de este conocimiento en tanto me sirva para resolver ese tipo de problemas' (Berger y Luckmann, 1966, pp. 61-62. Cursiva de los autores). Podría pensarse que el conocimiento de sentido común está compuesto por reglas y conocimiento proposicional no explicativo (información acerca de los objetos cotidianos de nuestro interés). Sin embargo, es razonable pensar que muchas de nuestras reglas

una persona ingiera, o compre, etcétera, esa bebida. Asignamos poder causal al contenido del estado intencional. Por otra parte, observemos que la afirmación 'el whisky causó que dijera eso' puede interpretarse de diferentes formas: o bien puede estar haciendo referencia a la eliminación de inhibiciones que genera el consumo de la sustancia, o bien puede hacer referencia al contenido específico de la elocución. No parece plausible la afirmación de que la sustancia ingerida *causó* el contenido específico de la elocución. Parece más razonable creer en la existencia de una cadena de estados mentales intencionales (creencias, deseos, etc.) que determina el contenido de aquella.

El segundo modo de comprensión psicológica analizado por Bieri enfrenta otra clase de dificultades. La cuestión crucial aquí es la posibilidad de adjudicar racionalidad a la conducta en el caso de que las razones no intervengan en el proceso productivo que conduce a aquella.

La relación entre razones y conducta ha sido objeto de un largo debate en las últimas décadas. Si seguimos en esta cuestión a Davidson (1963), el *explanans* de una explicación racionalizadora, esto es, las razones primarias de la acción (creencias y deseos), son a su vez *causas* de la acción. Una explicación de esta clase sería entonces, desde este punto de vista, racionalizadora *y* causal. Es bien conocido el hecho de que esta propuesta de Davidson ha sido ampliamente influyente en el rechazo de la tesis, dominante hasta ese momento, de que las razones no pueden ser causas, por lo cual una explicación racionalizadora no puede ser considerada una explicación causal.

Es verdad que la afirmación de que la conducta, para ser racional, debe ser *causada* por el contenido de las razones, puede ser cuestionable; podría quizás postularse otra clase de conexión entre razones y conducta que no fuese causal. Sin embargo, no parece razonable que podamos afirmar que la conducta pueda ser racional en el caso de que las razones no intervengan en absoluto en el proceso productivo que desencadena la conducta, esto es, no constituyan un factor eficaz (al menos parcial) en la producción de tal conducta. Si la única relación existente entre

---

cotidianas de acción (por ejemplo, llamar a la empresa telefónica cuando nuestro servicio se ve interrumpido) se basan en explicaciones tácitas, que podrían formularse explícitamente en caso necesario.

razones y conducta consiste en que aquellas son una clase de 'subproducto' de los procesos fisiológicos que causan efectivamente la segunda (o de procesos concomitantes a ellos), ¿cómo es posible predicar racionalidad de la conducta? Recordemos que ésta sería causada únicamente por procesos fisiológicos; las razones no intervendrían en absoluto en tal proceso productivo. Parece difícil afirmar que la conducta pueda ser racional dentro de este marco; sólo podrían ser racionales las creencias del supuesto agente o de un observador externo acerca de sus conductas.

La cuestión aquí no es, entonces, si las explicaciones racionalizadoras son o no extensibles: la cuestión es si es plausible considerar racional a la conducta en el caso de que las razones no hayan tenido poder productivo alguno en su aparición. En ausencia de una conexión más fuerte que la mera correlación entre razones y conducta, no parece que la racionalidad de ésta pueda afirmarse.

Una consideración final con respecto a la reformulación de las explicaciones psicológicas dentro de un marco epifenomenista se refiere al tipo de teoría de la explicación que tal reformulación requiere aceptar.

En caso de que el epifenomenismo fuese verdadero, cualquier programa fáctico de investigación de las causas de la conducta, razonablemente, debería estar destinado al logro de explicaciones de aquella en términos de las propiedades puramente físicas de los procesos internos. Con el desarrollo de la investigación, resultaría plausible esperar que se obtendrían explicaciones completas de la conducta en tales términos. En tal caso, se plantearía la cuestión de la coexistencia de explicaciones puramente físicas de la conducta con explicaciones psicológicas reevaluadas a la luz del marco epifenomenista. Dicho en otros términos, se plantearía el problema de la sobreabundancia de explicaciones. Por supuesto, se argumentará que tales explicaciones no compiten, lo cual es correcto ya que las explicaciones psicológicas no hacen referencia a los procesos que causan la conducta; sin embargo, debería probarse la necesidad de desarrollar, con fines cognoscitivos, las explicaciones psicológicas. En el mejor de los casos, Bieri habría probado (lo cual, como hemos visto, es dudoso) que las explicaciones psicológicas pueden tener sentido aun cuando se acepte el epifenomenismo; es algo muy distinto

probar la necesidad de desarrollar de una manera sistemática tal tipo de explicaciones.[74]

En segundo lugar, parece plausible pensar que las explicaciones psicológicas sólo podrían existir en el marco del epifenomenismo si se acepta la perspectiva que Kim denominaría 'internalismo explicativo'. Esto es así debido a que tales explicaciones no harían alusión a los procesos productivos que originarían la conducta.

Estas cuestiones exceden los problemas vinculados con el epifenomenismo y la causación mental; sin embargo, parece claro que es un problema adicional que el intento de preservar las explicaciones psicológicas dentro de un marco epifenomenista debe enfrentar, problema que afectaría a todos los tipos de explicaciones analizados por Bieri.

La posición de Bieri, en síntesis, no ofrece razones de peso para aceptar que haya logrado mitigar una de las consecuencias problemáticas del epifenomenismo, a saber, la irrelevancia explicativa de las propiedades mentales.[75]

El análisis ha estado restringido al plano de las explicaciones de sentido común; sin embargo, el epifenomenismo no sólo amenaza tales explicaciones, sino también las explicaciones científicas en psicología. Las dificultades que se encontrarán para admitir la relevancia explicativa de las propiedades mentales en el ámbito de la psicología científica serán de otro orden y, sostendremos, más complejas de resolver.

5. La solución de Jackson y Pettit: la explicación por programa

Como hemos observado en el capítulo I, Jackson y Pettit han planteado una dificultad aparentemente seria para la relevancia causal de las propiedades pertenecientes al dominio de las ciencias especiales y, en principio, para su relevancia explicativa. La consecuencia de la admisión conjunta de los cuatro

[74] Sobre todo si, como trataremos de mostrar luego, tales explicaciones no pueden constituir la base para la acción racional dirigida a la modificación de las conductas.

[75] Debe hacerse notar que este aspecto es más problemático en relación con las consecuencias gnoseológicas del epifenomenismo; en sí mismas, para nuestra concepción del mundo parecen mucho más graves otras consecuencias, como la autoalienación o el error masivo con respecto a nosotros mismos.

supuestos expuestos es que tales propiedades, que no tienen eficacia causal, carecerían de relevancia causal y explicativa. No obstante, Jackson y Pettit consideran que hay una manera de eludir estas consecuencias. Un segundo tipo de relación entre una determinada propiedad y un efecto sería aquella en la cual 'la realización de la propiedad asegura (…) que una propiedad productiva crucial sea realizada y que, en esas circunstancias, el suceso, bajo una cierta descripción, ocurra' (1990, p. 114).[76] Habría entonces al menos dos maneras distintas en las que una propiedad podría ser causalmente relevante: siendo eficaz en la producción de un determinado efecto, o 'programando' [*programming for*] la presencia de una propiedad eficaz. En este último caso, la ejemplificación de la propiedad no participaría en el proceso productivo que conduce al suceso, pero aseguraría que participe la ejemplificación de la propiedad que se requiere para tal proceso. De esta forma, la fragilidad del cristal aseguraría, por el mismo significado del término 'frágil', que el cristal tiene una estructura molecular que sería suficiente, dadas las circunstancias, para producir la rotura; análogamente, la temperatura del agua aproximadamente aseguraría que una molécula apropiadamente ubicada tendría un momento suficiente como para causar la fractura de la ligazón molecular en el recipiente y por lo tanto producir su rotura. Los autores caracterizan el rol de esta propiedad diciendo que su realización 'programa' para la aparición de la propiedad productiva y, bajo una cierta descripción, para el suceso producido. En sus propios términos:

> La solución propuesta para el problema que hemos estado afrontando es que, en cada caso, la propiedad de orden superior, ineficaz, es causalmente relevante para el efecto producido, porque su realización programa para la realización de una propiedad eficaz de orden inferior y, en las circunstancias, para la ocurrencia del suceso en cuestión (*ibid.*, p. 115).

La propuesta no implica que en todos los casos debamos ser capaces de identificar la propiedad causalmente eficaz a la que se haga referencia en la explicación por proceso; la 'explicación por programa' [*program explanation*] sólo proporcionará la base para creer que alguna de tales propiedades eficaces se encuentra en acción. Aun una escasa reflexión, observan Jackson y Pettit, sugiere

[76] 'The realization of the property ensures (…) that a crucial productive property is realized and, in the

que quizás la mayoría de las explicaciones que llegaremos a ofrecer serán explicaciones por programa.

Jackson y Pettit afirman además que la noción de 'propiedad programadora' [*programming property*] no sólo explica de que manera una propiedad causalmente ineficaz puede ser causalmente relevante, sino que muestra como una explicación por programa puede tener una significación que permanezca aun en presencia de una explicación 'por proceso' que involucre las propiedades eficaces correspondientes. Esto último se debe a que una explicación por programa de un suceso *e* puede proveer información que la correspondiente explicación por proceso no proporciona.[77] Aun cuando alguien comprenda los procesos de nivel inferior relevantes no necesariamente contará con la información disponible para quien tenga acceso a la explicación por programa.

Kim (1998) analiza la propuesta de Jackson y Pettit. Comienza señalando que tal propuesta es diferente, en un aspecto crucial, de otras alternativas de solución:[78] estos autores comienzan por admitir que las propiedades mentales, así como otras propiedades postuladas por las ciencias especiales y aun por ciertas ciencias naturales, carecen de eficacia causal. Admitida esta premisa, queda por considerar si puede salvarse su *relevancia* causal.

El enfoque de las explicaciones 'por programa', observa Kim, no difiere sustancialmente de su modelo de causación superveniente. En efecto, la idea principal de este modelo es que 'una propiedad puede derivar su rol causal y tener un rol en una explicación causal, en virtud de su superveniencia sobre una propiedad involucrada en procesos causales' (p. 74). Asimismo, en el modelo de Jackson y Pettit, la presencia de la propiedad *F* 'asegura' o 'programa para' la ocurrencia de una propiedad de base causalmente eficaz.

Dado que Jackson y Pettit comienzan admitiendo que las propiedades mentales, así como las propiedades de las ciencias especiales, carecen de poderes causales, observa Kim, parece justo considerar su posición como epifenomenista. El

---

circumstances, that the event, under a certain description, occurs.'

[77] Si bien, observan Jackson y Pettit, hay excepciones, como el caso de la fragilidad.

[78] Como las de Baker, Burge y otros (*supra*, p. 128).

modelo de la explicación por programa parece completamente compatible con el epifenomenismo: el epifenomenista puede admitir perfectamente que la ocurrencia de un dolor 'asegura' o 'programa para' la presencia de su causa neural, la activación de las fibras C, y este suceso causa en la persona el dolor.

Kim se pregunta, razonablemente en nuestra opinión, si, en ausencia de eficacia causal de las propiedades mentales, es apropiado hablar de relevancia causal. A su modo de ver, parece difícil que quede espacio para la relevancia causal; la única relevancia de la cual le parece apropiado hablar es la *relevancia informacional*: la ocurrencia de una propiedad programadora daría información acerca de que cierta propiedad causalmente eficaz está presente y en acción, si bien puede ignorarse cuál es.

La pregunta que plantea Kim con respecto a este enfoque es la siguiente: ¿tiene éxito el modelo de la explicación por programa en reivindicar las explicaciones de las ciencias especiales y en demostrar la eficacia explicativa de las propiedades de tales ciencias? La respuesta, observa, depende de cual sea el punto de vista que de adopte con respecto a la explicación. Si se está dispuesto a vivir con un modelo más débil y laxo de relevancia explicativa, renunciando a la causación mental y a la explicación causal mentalista, puede recurrirse a la idea postulada por David Lewis referente a que explicar un suceso es proveer cierta información acerca de su historia causal. Sin embargo, debe efectuarse a este principio una modificación a la que considera 'no insustancial': en la concepción de Lewis la historia causal de un suceso incluye a éste y a cualquiera de sus partes y está cerrada bajo la relación de dependencia causal. Sin embargo, las historias causales no están cerradas bajo la relación conversa de la causación: los epifenómenos no son parte de la historia causal del suceso. Esto implica que, bajo el modelo de Lewis, invocar un epifenómeno (por ejemplo, el dolor) de una causa real (la activación neuronal) del suceso a ser explicado no constituirá una explicación, o una explicación causal, de ese suceso. La modificación que debe efectuarse es, entonces, sostener que la red causal de un suceso está cerrada bajo la relación de dependencia causal y su conversa, y luego caracterizar la idea de explicación en términos de una provisión de información acerca de la red causal en la cual el suceso en cuestión está inserto.

Kim observa que sólo este tipo de noción extremadamente laxa de explicación puede hacer lugar a la explicación por programa propuesta por Jackson y Pettit. Sin embargo, no considera que haya razón para rechazar tal noción laxa de explicación:

> La noción de explicación es muy laxa y flexible –esencialmente tan laxa y flexible como las nociones subyacentes de comprensión y de hacer algo inteligible- y nadie debería legislar acerca de que cuenta como explicación, excepto sólo que cuando hablamos de 'explicación causal' deberíamos insistir, como he dicho, en que lo que se invoca como una causa sea realmente una causa de aquello que está siendo explicado. El realismo acerca de la explicación debería al menos cubrir la explicación causal (p. 76).

Kim considera entonces que la pregunta principal acerca de la explicación por programa no es la referente a la noción de explicación en juego, sino la referente a si esta clase de explicación puede reivindicar las propiedades invocadas por las ciencias especiales, en particular las propiedades psicológicas. La respuesta, considera, es que tal manera de salvar la relevancia explicativa es demasiado débil para ser satisfactoria. En su opinión, cualquier reivindicación de la explicación psicológica debe hacer justicia al 'porque' en afirmaciones tales como 'ella dio un respingo porque sintió un dolor repentino en su codo', y hacer esto requiere de un sentido más robusto del 'porque' que el que es provisto por la explicación por programa.

5.1. Explicación por programa y efectividad instrumental

Coincidimos con Kim en que el enfoque epifenomenista de Jackson y Pettit no hace justicia a la relevancia causal de propiedades psicológicas. Sin embargo, el modelo de la explicación por programa presenta adicionalmente, en nuestra opinión, una limitación que disminuye en una medida considerable su atractivo. En lo sucesivo, nos centraremos en el caso de las explicaciones psicológicas; no obstante, mucho de lo que se dirá puede ser igualmente válido, *mutatis mutandis*, en el ámbito de otras ciencias especiales.[79]

[79] Al desarrollar este argumento estaremos admitiendo varios presupuestos que conviene explicitar. En primer lugar, una concepción realista de la ley científica, que constituye la base para la explicación y la acción. Las leyes teóricas (o al menos una parte no menor de ellas) deben hacer referencia tanto a entidades y propiedades realmente existentes como a relaciones objetivas que conectan esas entidades y

Comencemos aplicando el argumento de Jackson y Pettit al caso de una acción intencional del tipo usualmente explicado por la psicología. Supongamos que una persona manifiesta verbalmente su rechazo a compartir su ámbito de trabajo con alguien perteneciente a una minoría étnica. Siguiendo el modelo de Jackson y Pettit,[80] dos explicaciones de la conducta serían posibles: a) porque la persona tiene fuertes prejuicios en contra de esa minoría étnica y, por extensión, contra cualquier representante de ese grupo; b) por la existencia de una determinada configuración neuronal que precede (causa) la acción. Siguiendo ahora el argumento, la explicación a) no puede hacer referencia a una propiedad causalmente eficaz (la disposición psicológica), ya que i) la disposición ha sido eficaz sólo si la configuración neuronal lo ha sido (satisface 3a); ii) la disposición no colabora en la producción de la configuración neuronal (satisface 3b); por último, la disposición y la configuración neuronal no constituyen factores causalmente coordinados (satisface 3c).

Este tipo de ejemplos muestra cómo el argumento de Jackson y Pettit excluye la posibilidad de que, en caso de que dispongamos de dos explicaciones de una conducta, una basada en una propiedad psicológica y otra basada en una propiedad neurofisiológica, la explicación basada en la propiedad psicológica sea aceptable, ya que tal propiedad no resultaría causalmente eficaz, mientras que la otra sí lo sería. No puede existir, por lo tanto, una explicación causal apropiada de un suceso físico (la conducta) en términos de un suceso mental (causación de lo mental a lo físico). Una extensión del argumento muestra cómo tampoco pueden darse casos de causación de estados mentales por otros estados mentales. Examinemos el siguiente ejemplo: una persona decide abandonar su carrera universitaria. Dos explicaciones son nuevamente posibles: a) la persona ha evaluado negativamente su rendimiento académico y considera que sus fracasos se deben a su falta de

---

propiedades. En segundo lugar, una concepción realista de la explicación, esto es, una concepción que supone que el concepto de explicación satisfactoria de hechos singulares incluye como condición necesaria (aunque no suficiente) el que la explicación esté fundamentada en alguna clase de relación de dependencia objetiva entre los fenómenos. Dado que este supuesto es aceptado por los autores que examinaremos a continuación, aceptarlo sin discusión parece perfectamente legítimo. Por último, la noción de que las reglas de acción tecnológicas se basan en leyes y explicaciones, en contextos conceptuales aplicados. (Para un análisis de la relación entre ley, explicación y regla, y sus respectivos contextos, *cfr*. Quintanilla (1988)).

[80] *Supra*, p. 39.

condiciones para esa disciplina;[81] b) por la existencia de un determinado proceso neurofisiológico subyacente. Una vez más, i) la creencia y la atribución han sido eficaces sólo si la configuración neuronal lo ha sido (satisface 3a); ii) la creencia y la atribución no colaboran en la producción de la configuración neuronal (satisface 3b); por último, la creencia, la atribución y la configuración neuronal no constituyen factores causalmente coordinados (satisface 3c). La generalización de argumentos de esta clase, entonces, parece probar que una propiedad mental no puede ser causalmente eficaz en la producción de otra propiedad mental.[82]

Si las explicaciones psicológicas del tipo (a)) fuesen correctas, según el enfoque de la psicología científica podríamos modificar la conducta o las creencias a partir de la modificación de la disposición o de las creencias y las atribuciones. Sin embargo, si los efectos (la conducta o la creencia) no son causadas por la disposición o por las creencias y las atribuciones, sino que son determinadas por las configuraciones o procesos neuronales subyacentes, este camino parece vedado. Los estados o procesos mentales antecedentes podrían ser funcionales para la modificación de las conductas o de otros estados mentales en tanto actuar sobre ellos nos permitiera modificar la configuración neuronal subyacente, pero esto no puede ser posible. Constituiría un caso de causación descendente, que, entre otras razones, violaría la clausura causal del mundo físico.[83]

La explicación por programa, entonces, haría referencia a propiedades causalmente ineficaces (en nuestros ejemplos, la disposición o las creencias), pero, como se ha dicho, cuya presencia 'programaría' la realización de una propiedad de orden inferior eficaz (las configuraciones neuronales) que serían causalmente responsables de la ocurrencia de los sucesos en cuestión. La presencia de las propiedades de orden superior, causalmente relevantes, 'aseguraría' la presencia de las propiedades de orden inferior, causalmente eficaces. De esta manera, podría

[81] La elección de los ejemplos no es azarosa. Se los ha elegido por pertenecer a teorías que plantean la determinación de la conducta o las creencias a través de otras creencias, disposiciones, actitudes e intenciones, y que tienen aplicaciones exitosas en la modificación de la conducta a través de la modificación de los determinantes antecedentes.

[82] Es pertinente observar que si el argumento de Jackson y Pettit afecta a todas las propiedades de las ciencias especiales, la explicación neuronal también se verá en peligro. Lo planteado tiene validez sólo a los fines del desarrollo del argumento que lleva a concluir la ineficacia causal completa de lo mental.

[83] Esta restricción no está presente explícitamente en Jackson y Pettit, pero parece un criterio perfectamente aceptable en este caso. *Cfr.* al respecto Sabatés (2001).

salvarse a las explicaciones psicológicas, análogamente a la manera en que se salvarían las explicaciones de las demás ciencias especiales.

Sin embargo, salvar las explicaciones de las ciencias especiales implica salvar las propias ciencias especiales; dicho en otros términos, salvar las explicaciones de las ciencias especiales implica salvar los principios explicativos, y estos principios no son otros que los enunciados teóricos de las ciencias especiales. Pero las ciencias especiales (como cualquier otra ciencia) no sólo tienen intereses cognoscitivos, sino también intereses extracognoscitivos: fundamentalmente, la modificación racional del mundo.[84] Por lo tanto, la evaluación del modelo de la explicación por programa, entendemos, no puede estar limitada al análisis del plano puramente teórico. La acción racional presupone teorías explicativas verdaderas (o aproximaciones suficientemente buenas a éstas) que hagan referencia a las propiedades causalmente relevantes en la producción de los sucesos. Sólo en este caso puede esperarse que de esas teorías puedan derivarse tecnologías eficientes y planes de acción práctica racionales.[85] Sin embargo, dado que las explicaciones por programa no hacen referencia a las propiedades causalmente eficaces en la producción de un efecto, esto parece implicar, en principio, que no podemos valernos de ellas para la modificación racional del mundo; no podemos esperar que, por ejemplo, modificando ciertas propiedades psicológicas se vean modificadas otras propiedades psicológicas.

Jackson y Pettit no analizan el problema de la relación de la explicación por programa y de las propiedades 'aseguradoras' con la eficacia de la acción instrumental, con la excepción del comentario que sigue:

> [S]in importar cómo sea entendida la noción de eficacia causal, es distinta de la noción de efectividad instrumental. Una propiedad contará como instrumentalmente efectiva *vis-à-vis* un efecto particular si realizar esa propiedad hubiera sido una buena táctica para la producción del efecto. Pero tal efectividad no implica eficacia: no significa que el efecto ocurrió en virtud de la ejemplificación de la propiedad (p. 109).

[84] Dejamos de lado aquí, por ser tangenciales a los propósitos de este análisis, consideraciones relativas a la distinción entre ciencia, tecnología y praxiología, y a sus respectivos intereses cognoscitivos y prácticos.

[85] Mario Bunge es uno de los autores que ha enfatizado reiteradamente la importancia de la utilización de teorías explicativas verdaderas como base para la acción racional guiada por la tecnología. *Cfr*. su (1983), en especial el capítulo 11.

La manera de interpretar estas afirmaciones, en nuestra opinión, no resulta evidente. Una manera de entenderlas es suponer que del hecho de que una propiedad sea instrumentalmente efectiva no puede inferirse que sea causalmente eficaz (opción gnoseológica), mientras que una segunda manera sería sostener que *hay* otras formas, además de la eficacia causal, en que una propiedad pueda resultar instrumentalmente efectiva (opción ontológica). Ahora bien, si cuando en la tercera oración se hace referencia a la 'eficacia', se está hablando de eficacia causal, la afirmación parece plausible. Por ejemplo, parece posible producir la ejemplificación de una propiedad mediante la manipulación de su base de superveniencia. Pero si cuando se hace referencia a la eficacia, se está hablando de eficacia a secas, entonces sus consecuencias son en verdad notables. Supongamos una propiedad que es producida por otra, ya sea a través de una relación causal, o por superveniencia, o por medio de alguna otra clase de relación de dependencia objetiva, pero que en sí misma es completamente impotente: no causa la existencia de ninguna otra propiedad, ni a partir de ella sobrevienen propiedades de nivel superior, ni incide en absoluto en ninguna clase de proceso productivo. Al margen de que pueda resultar razonable dudar de su existencia,[86] no parece posible que una propiedad absolutamente carente de poder productivo pueda resultar instrumentalmente efectiva en modo alguno. Adoptemos entonces la primera interpretación, que parece más razonable, y examinemos sus consecuencias.

Las posibilidades son, entonces, las siguientes: una propiedad es instrumentalmente efectiva porque es causalmente eficaz, o porque es una buena táctica para la producción de otra propiedad, sin ser causalmente eficaz. Si la propiedad es epifenoménica (como Jackson y Pettit admiten que son las propiedades de las ciencias especiales), debemos pensar bajo qué condiciones puede darse la segunda alternativa. Puede ejemplificarse fácilmente esta posibilidad con una acción técnicamente a nuestro alcance: modificando la dotación genética de un organismo (genotipo –propiedad de nivel inferior-)[87] logramos la modificación de sus características físicas observables (fenotipo –propiedad de nivel superior-). Podemos, entonces, suponer que una manera de intervenir en los fenómenos

[86] Por supuesto, esto dependerá del criterio de realidad que se acepte.

instrumentalmente efectiva sin que exista eficacia causal es actuando sobre las propiedades de base (subvenientes) para lograr efectos no causales sobre las propiedades sobrevinientes. [88]

Sin embargo, esta posibilidad no parece constituir una réplica apropiada a la objeción que hemos expuesto. Consideremos una vez más las propiedades psicológicas. Supongamos en primer lugar la existencia de propiedades de orden superior a las propiedades psicológicas, esto es, propiedades sociales. Según el esquema expuesto, parecería que podríamos influir en las propiedades sociales a través de la modificación de las propiedades psicológicas. Por ejemplo, modificando ciertas actitudes (propiedad psicológica –de nivel inferior) de los miembros de un grupo hacia éste puede lograrse un incremento de la cohesión grupal (propiedad social –de nivel superior). Pero para modificar las propiedades psicológicas pertinentes para la modificación de las propiedades sociales, no puede recurrirse a su vez a propiedades también psicológicas; esto implicaría, en principio, una acción causal de unas propiedades psicológicas sobre otras propiedades psicológicas, y sabemos que esta posibilidad está vedada, ya que la extensión del argumento de Jackson y Pettit impide admitir la causación de un estado mental por otro estado mental.[89]

Por lo tanto, si el modelo de Jackson y Pettit únicamente admite que propiedades de un cierto nivel sólo puedan ser realizadas o sobrevenir a partir de propiedades de nivel superior, pero que los poderes causales quedan reservados exclusivamente para las propiedades del nivel de la base (si es que este nivel existe), entonces no parece existir una forma en que podamos disponer de los cambios o

[87] Carece de importancia el hecho de que la manipulación del genotipo sea causal: una vez producida, las propiedades de nivel superior sobrevienen a partir de las propiedades (manipuladas) de nivel inferior.

[88] Estamos admitiendo aquí que las relaciones que pueden fundamentar la efectividad instrumental sólo pueden ser relaciones objetivas del tipo de la causación o la superveniencia, no relaciones de tipo conceptual como la dependencia de Cambridge.

[89] Es verdad que si pudiéramos producir propiedades psicológicas invariablemente acompañadas por otras propiedades psicológicas sobre las cuales las primeras carecen totalmente de eficacia causal, debiéndose esta correlación a la existencia de mecanismos de nivel inferior, sería correcto afirmar que al modificar ciertas propiedades psicológicas se verían modificadas otras propiedades psicológicas. De esta manera, la ineficacia causal no se vería acompañada de inefectividad instrumental. Sin embargo, esto no parece posible en la práctica dentro del marco de Jackson y Pettit. Como hemos señalado, si las propiedades psicológicas carecen completamente de efectos causales (tanto en su propio nivel cuanto en niveles inferiores) no es posible que ejerzan una influencia de ninguna clase sobre los mecanismos que median en la correlación, y que son los responsables de la ejemplificación de la segunda clase de propiedades psicológicas.

manipulaciones en las propiedades de un nivel con el fin de producir modificaciones en otras propiedades del mismo nivel. Y, como consecuencia, las explicaciones por programa no pueden hacer justicia a un rasgo característico de la actividad científica: basar la acción eficaz en explicaciones verdaderas.[90]

Tal vez las explicaciones por programa podrían tener un interés instrumental indirecto en el caso de que proveyeran información acerca de las propiedades causalmente eficaces que están actuando en ese momento, cuya identificación permitiría construir una explicación por proceso. Sin embargo, hemos visto que no lo hacen: Jackson y Pettit sostienen que proporcionar una explicación por programa solamente proporciona las bases para creer que alguna propiedad eficaz está actuando.

Quisiéramos agregar aquí dos observaciones acerca de la objeción expuesta.

Para que la objeción tenga fuerza no es necesario que se recurra a una teoría causalista de la explicación (esto es, la concepción según la cual no hay explicaciones no causales de sucesos singulares). Esta idea, defendida por algunos teóricos de la explicación (por ejemplo Salmon (1984)), es rechazada por otros (Hempel (1965), Kim (1974) y Ruben (1990), entre otros). Estos autores defienden la idea del 'pluralismo explicativo' (denominación que tomamos de Sabatés (1996)), esto es, la idea de que hay explicaciones no causales de sucesos singulares. Sólo es necesario, como hemos señalado, un principio de realismo explicativo, que sostenga que las explicaciones de sucesos singulares deben estar fundamentadas en alguna relación objetiva de dependencia o determinación. Tampoco es necesaria la idea de que *todas* las explicaciones de una disciplina como la psicología tienen que hacer referencia a propiedades con capacidad productiva. Por ejemplo, puede admitirse perfectamente que una explicación de leyes por subsunción bajo ciertos principios teóricos de mayor generalidad constituye una explicación de la ley perfectamente aceptable, pese a que no se base en ningún principio causal.[91] Es suficiente con que algunas

[90] Por supuesto, la acción racional podría ser ineficaz aun cuando las propiedades a las que se apela posean poderes causales; sin embargo, el que las propiedades posean capacidad productiva parece condición *sine qua non* para poder actuar de manera eficaz.

[91] *Cfr.* al respecto Hempel (1965). Como se observa en ocasiones en los textos de filosofía de la ciencia, la investigación científica está primariamente dirigida hacia la búsqueda de explicaciones de regularidades, y no de hechos singulares. La explicación de regularidades presenta problemas extraordinariamente complejos que no podemos abordar aquí, pero parece necesario poner de manifiesto

explicaciones lo sean, ya que Jackson y Pettit, por su propio argumento, tienen que negar que cualquier propiedad psicológica pueda ser productiva. Vastas áreas de la psicología intencional (entre ellas, de la psicología cognitiva y social) se verían afectadas de manera decisiva en caso de que el epifenomenismo fuese verdadero.

Coincidimos con Kim en la idea de que la noción de la explicación es laxa y flexible; la explicación científica es un tema sobre el que dista de haber consenso. Sin embargo, una propuesta que pretenda salvar a las explicaciones de las ciencias especiales debería, en principio, ser capaz de preservar ciertas características básicas que podemos razonablemente atribuir a tales explicaciones.[92]

Por último, conviene recordar que el argumento expuesto en contra del modelo de la explicación por programa no es un argumento en contra del epifenomenismo en general, sino un argumento en contra de una posición que combine una forma de epifenomenismo con el intento de salvar las explicaciones psicológicas en tanto que ciencia especial. No todo epifenomenista tiene necesariamente que estar interesado en salvar a las ciencias especiales; un epifenomenista podría perfectamente, sin contradicción, admitir la ineficacia causal de las propiedades de cierto nivel y extraer como consecuencia la necesidad de renunciar a las explicaciones que hagan referencia a tales propiedades.

6. El rol explicativo de lo mental en un dualismo aislacionista de propiedades

Sabatés (1997) aborda en profundidad la cuestión de si la ineficacia causal acarrea inevitablemente la irrelevancia explicativa. En particular, se pregunta (pregunta que resulta especialmente pertinente a los fines de nuestro análisis) si los psicólogos cognitivos deberían estar preocupados por esta posibilidad. De manera preliminar, si se considera que las explicaciones deben reflejar las relaciones causales existentes entre los sucesos a los que se hace referencia en la explicación, y los

---

que tales explicaciones no parecen ajustarse al modelo de la explicación basada en una relación objetiva, al menos en lo que respecta a la relación causal.

[92] Por supuesto, no toda la búsqueda de explicaciones en psicología (como en las restantes ciencias) está guiada por objetivos prácticos. Si tenemos en cuenta la fatigada distinción entre ciencia básica y aplicada, debemos admitir que, al menos en una proporción no marginal de casos, la inaplicabilidad de las explicaciones psicológicas con objetivos prácticos no resultaría problemática; la búsqueda de explicaciones constituiría un fin en sí mismo en la ciencia básica.

argumentos tendientes a mostrar la ineficacia causal de la mente son correctos, lo mental parecería no tener lugar en las explicaciones de la conducta. Esta consecuencia es evidentemente indeseable no sólo para la psicología cognitiva, sino para toda psicología científica, ya que no parece razonable que ninguna ciencia especial pueda sobrevivir si no puede proporcionar explicaciones propias. Su conclusión es que, *prima facie*, los psicólogos cognitivos deberían estar preocupados por la ineficacia causal de lo mental.

Una actitud resignada ante el argumento de la exclusión es, en opinión de Sabatés, aceptar el epifenomenismo. Como hemos visto en este capítulo, el epifenomenista resueltamente admite que lo mental carece de poderes causales, pero permite preservar en apariencia los restantes rasgos centrales del materialismo no reduccionista. Sin embargo, Sabatés considera que una extensión del argumento de la exclusión permite descartar al epifenomenismo como alternativa: las propiedades mentales no serían efectos de propiedades físicas. La solución que favorece es un dualismo de propiedades que acepte el aislamiento causal de lo mental (en lo sucesivo, 'aislacionismo' [*isolationism*]). Según este enfoque, las propiedades mentales sobrevendrían mereológicamente como propiedades de organismos totales; existirían paralelamente con sus bases físicas causalmente cerradas, sin el más mínimo contacto causal, de acuerdo con una 'armonía preestablecida' dada por una relación de superveniencia que estructuraría sincrónicamente la realidad. Este punto de vista, en su opinión, permite mantener varias de las tesis centrales del materialismo no reduccionista (o 'dualismo de propiedades', como prefiere denominarlo). Este 'dualismo aislacionista de propiedades' debe resolver dos problemas centrales: determinar un criterio apropiado de realidad de las propiedades (para evitar el irrealismo de lo mental), y proponer modelos de explicación en los cuales jueguen un rol las propiedades mentales causalmente inertes (para evitar la irrelevancia explicativa).[93] Expondremos a continuación su respuesta a esta última cuestión.

Una concepción realista de la explicación, observa Sabatés, no sólo requiere de relaciones objetivas que fundamenten las explicaciones; la explicación es una

[93] Para los argumentos que conducen desde la ineficacia causal hasta la irrelevancia explicativa, véase el capítulo II.

empresa epistemológica, y lo que es considerado una buena explicación dependerá de la situación epistémica de aquél que la requiera. Ahora bien, el afirmar que la explicación debe estar basada en relaciones objetivas no significa afirmar que la única relación capaz de cumplir este rol debe ser la relación causal; otras relaciones pueden ser apropiadas para este objetivo. Entre ellas, la superveniencia ha sido la más profundamente estudiada. La admisión de que existen relaciones objetivas de dependencia no causales le permite formular un principio de pluralismo explicativo:

[DC] C explica E sólo en caso de que C denote c y E denote e, e dependa de c, y donde C y E son entidades lingüísticas y c y e son sucesos o propiedades.

Este principio permite liberarse de las restricciones impuestas por el causalismo explicativo, y por lo tanto bloquear el argumento que conduce de la ineficacia causal a la irrelevancia explicativa. Este resultado, considera, conjuntamente con el intento de defender la realidad de las propiedades mentales, es muy importante; de esta forma, un dualismo aislacionista de propiedades, si bien admite la ineficacia causal de las propiedades mentales, puede mantener muchos de los rasgos importantes de las versiones más convencionales. Sin embargo, queda la tarea de mostrar de qué manera las propiedades mentales causalmente aisladas pueden jugar un rol de importancia en las explicaciones.

Tres son los casos de posibles explicaciones que involucren propiedades mentales: la explicación de propiedades mentales en términos de propiedades físicas (característicamente neurales), la explicación de propiedades físicas (característicamente conductuales) en términos de propiedades mentales, y la explicación de propiedades mentales en términos de otras propiedades mentales. Siguiendo la terminología introducida por Cummins (1983), Sabatés denomina 'explicaciones de transición' a aquellas explicaciones que intentan explicar cambios en estados o propiedades de un objeto u organismo como efecto de estados o propiedades previas, y 'explicaciones no transicionales de propiedades' a aquellas explicaciones que nos ayudan a comprender cómo un objeto o un organismo ejemplifica una determinada propiedad. Podrán considerarse explicaciones de

transición a todas aquellas explicaciones en las cuales el *explanans* es temporalmente previo al *explanandum*.

Valiéndose de estos conceptos, presenta tres casos de explicaciones que involucran propiedades mentales. El primer caso será aquel en el cual una propiedad mental es explicada en términos de su base de superveniencia (la propiedad física a partir de la cual sobreviene). Por ejemplo, la percepción de un cuadrado rojo será explicada en base a un estado neural complejo que constituye su base de superveniencia. Esta clase de explicación, en su opinión, es tan perfectamente aceptable para un realista explicativo como una explicación basada en una relación causal.

Pueden afirmarse entonces dos obvios principios que gobiernan los casos de explicaciones de explicaciones causales de transición, así como explicaciones de propiedades:

0. Cuando X causa Y, en condiciones epistémicas apropiadas Y explica transicionalmente a X.
1. Cuando X sobreviene mereológicamente a partir de Y, en condiciones epistémicas apropiadas Y explica no transicionalmente a X.

Una variación del caso 1, observa Sabatés, es aquella en la cual proponemos una explicación de M en términos de la causa de P, denominada P-. por ejemplo, explicaríamos nuestro estado perceptivo actual en términos de una determinada condición física en nuestro entorno. Esta condición física (P-) es la causa de nuestro estado neural actual (P), que a su vez es la base de superveniencia de nuestro estado perceptivo (M). Esta clase de explicación es perfectamente aceptable: un realista pluralista, señala, estaría feliz de apoyar explicaciones de este tipo, en las que hay una dirección de dependencia que conecta el *explanans* con el *explanandum*. Explicaciones de esta clase son denominadas 'explicaciones impuras de transición' [*impure transition explanations*]. Si bien algunos teóricos han defendido la idea de que las explicaciones de transición sólo pueden ser causales, es posible, en su opinión, preservar la intuición que subyace a este punto de vista sin adoptar éste en sí mismo. El que estemos forzados a aceptar que detrás de toda explicación de transición debe haber un proceso causal subyacente no implica la existencia de un proceso causal que

conecte el suceso o propiedad a ser explicado con aquel que lo explica. Esto autoriza a aceptar las explicaciones de transición impuras. Esta posibilidad conduce al siguiente principio explicativo adicional:

1'. Cuando X causa Y, y Z sobreviene mereológicamente a partir de Y, en condiciones epistémicas apropiadas X explica (impuramente) transicionalmente a Z.

Examinemos el segundo caso planteado por Sabatés. Supóngase que se desea explicar una ocurrencia conductual (P). Existirá una causa de esta conducta (P-), en términos de determinado suceso neural, que podrá constituir una explicación apropiada de la acción. Sin embargo, si requerimos de una explicación al nivel de las propiedades mentales, podremos construir tal explicación en términos de deseos y creencias que anteceden a la acción. Podría ocurrir, por otra parte, que la explicación de la conducta en términos neurales no fuera accesible para nosotros. Hay razones para pensar que la explicación en términos de deseo-creencia sería preferible, pero se presenta el obvio problema de que tal explicación es una explicación de transición que presupone la causalidad de lo mental a lo físico (vedada por los argumentos de la exclusión). Pero a partir de la variación del caso 1 puede decirse que estamos autorizados a aceptar explicaciones impuras de transición, siempre que se cumplan los requisitos que reclama el realista explicativo.

Por esta razón, Sabatés considera que está justificado adoptar el siguiente principio explicativo:

2. En los casos en que X causa Y, y Z sobreviene mereológicamente a partir de X, en condiciones epistémicas apropiadas Z explica transicionalmente a Y.

El tercer caso que analiza es el siguiente. Supóngase que tenemos una percepción de un tigre (M) y subsecuentemente una creencia de que nuestra integridad física está en peligro (M-), y queremos explicar la última en términos de la primera. También en este caso hay relaciones objetivas de dependencia entre los estados mentales y sus bases de superveniencia (los estados neurales), y para poder

proveer de una explicación de un suceso mental a otro suceso mental se necesita el siguiente principio:

3. En los casos en que X causa Y, Z sobreviene mereológicamente a partir de X, y W sobreviene mereológicamente a partir de Y, en condiciones epistémicas apropiadas Z explica transicionalmente a W.

Sabatés considera que los principios 2 y 3 son mucho más difíciles de aceptar que 1'. Si se compara 1' y 2, observa, puede notarse que ambos están basados en relaciones de dependencia, por lo que, en este respecto, no habría razones *a priori* para descartar a 2 como un tipo de explicación no causal de transición. Sin embargo, difieren en un aspecto importante: mientras que en 1' la dirección de la dependencia lleva desde el *explanandum* al *explanans*, esto no ocurre con 2. La relación de superveniencia en el caso 2 (asociado con el principio 2) parece ir en una dirección incorrecta, de P- a M-, pero resultaría necesario que la relación fuese de M- a P- para lograr una relación de dependencia compleja desde M- a P. En otros términos, sintetiza, mientras que en el caso 1a, M, indirectamente, depende de P-, en el caso 2 P no depende de M-. Lo que se requiere para mantener este último tipo de explicaciones es mostrar que una relación de dependencia entre *explanans* y *explanandum* no es necesaria para que el *explanandum* dependa del *explanans*.

La forma en que se puede lograr este objetivo, en opinión de Sabatés, es recurriendo a la relación de aseguramiento propuesta por Jackson y Pettit, adecuadamente suplementada.[94] Esta necesidad de suplementación se hace necesaria para evitar ciertos aspectos controvertidos.[95]

Una primera modificación que, en opinión de Sabatés, es necesario realizar a la noción de aseguramiento es permitir que se ajuste al pluralismo explicativo. En efecto, Jackson y Pettit sólo tienen en cuenta, en su modelo, relaciones causales. Como hemos observado, consideran que explicar algo es proporcionar información acerca de su historia causal. Pero si se admiten, como hace Sabatés, explicaciones no

[94] Sabatés rechaza la posibilidad de que un enfoque basado en los contrafácticos pueda lograr este objetivo.

[95] Uno de estos aspectos es el obvio contraejemplo proporcionado por el caso de la relación entre la propiedad aseguradora y la propiedad asegurada. Este contraejemplo manifiesta en casos como los siguientes: S tiene la propiedad neural P porque tiene un dolor agudo; las partes a, b y c de una totalidad w tienen las propiedades P, Q y R respectivamente porque w tiene la propiedad T.

causales, la noción de aseguramiento debe ajustarse adecuadamente.[96] Debe ser posible aceptar propiedades aseguradoras que puedan ser explicativamente relevantes cuando estén combinadas con relaciones no causales. Un ejemplo de una tal explicación sería el siguiente: 'esta pieza de metal tiene tal conductividad térmica debido a que tiene tal conductividad eléctrica'. Si concedemos que ésta es una explicación aceptable, sostiene Sabatés, podemos decir que una conductividad eléctrica dada asegura cierta estructura molecular (no una particular), la cual es necesaria para una determinada conductividad térmica.

Una segunda modificación que es necesario efectuar a la noción de aseguramiento para descartar resultados indeseables es requerir que la relación aseguradora no pueda ser resultado de relaciones meramente estipulativas o semánticas, y que en general deba ser lógica o conceptualmente contingente el hecho de que la propiedad aseguradora o de nivel superior asegure a la propiedad de base. Este requisito permite descartar 'explicaciones' indeseables del tipo 'tuvo lugar un funeral en Atenas esta mañana porque Jantipa enviudó'.

La tercera y última modificación necesaria es el requisito de que, para ser explicativa, la propiedad aseguradora no puede ser temporalmente posterior a la propiedad asegurada y, en general, debe ser sincrónica con la propiedad asegurada o 'fuente'. Esta condición permite descartar explicaciones como la siguiente: 'este trozo de papel está ardiendo debido a la rojez del carbón'.

Los tres requisitos que deben exigirse para que una relación de aseguramiento sea explicativa son, en síntesis: a) que esté ligada a relaciones que ya sean explicativas; b) que sea lógicamente contingente; y c) que sea sincrónica.

La ineficacia causal, concluye Sabatés, no implica entonces irrelevancia explicativa. Retomando su pregunta inicial (aunque no lo dice explícitamente), podría agregarse que los psicólogos (cognitivos o no) no deberían estar preocupados por la ineficacia causal.

[96] Al señalar que Jackson y Pettit sostienen que las propiedades mentales pueden ser causalmente relevantes a pesar de ser causalmente ineficaces, Sabatés considera que su posición merece ser llamada 'cuasiepifenomenismo'.

El intento de Sabatés representa, en nuestra opinión, una mejora con respecto al modelo de la explicación por programa de Jackson y Pettit, en particular en lo que se refiere al ajuste al pluralismo explicativo.[97]

Sin embargo, resulta dudoso que no presente una limitación análoga a la que hemos observado en el modelo de estos autores.[98] En particular, si tenemos en cuenta la cuestión básica que parece originar el estudio, esto es, si los psicólogos (cognitivos) deberían estar preocupados por la ineficacia causal de lo mental. Las propiedades mentales, en su propuesta, están causalmente aisladas; en el mejor de los casos sobrevienen, a partir de ellas, propiedades de nivel superior, pero no podríamos servirnos de ellas para modificar otras propiedades psicológicas. Examinemos el principio 2. Z, admitiendo la corrección del argumento, es capaz de explicar transicionalmente a Y; sin embargo, no afecta, causalmente o de alguna otra forma, a X e Y. Lo mismo ocurre con los casos abarcados por el principio 3. En casos como éstos, Z y W (propiedades mentales) que sobrevienen mereológicamente a partir de X e Y respectivamente, no son, en principio, capaces de causar nada (por el argumento de la exclusión), pero tampoco parece posible que posean efectividad instrumental alguna. En el mejor de los casos, quizás podría imaginarse una situación en la cual (suponiendo que el argumento de la exclusión no sea generalizable), podríamos servirnos de las propiedades mentales para modificar propiedades de nivel superior –presumiblemente, propiedades sociales-. Pero esto no evita que sean impotentes no sólo para producir modificaciones en el nivel inferior (mundo físico), sino también en su propio nivel.[99]

En este capítulo hemos analizado algunas relaciones de las doctrinas que aceptan que las propiedades mentales carecen de poderes causales con los

[97] Veremos en el capítulo VIII que parece haber buenas razones para adherir a la tesis de que el realista explicativo no necesita adherir a la tesis del causalismo explicativo.

[98] Puede parecer que la objeción que presentamos se basa en una estrategia ilegítima: fijamos condiciones de adecuación de las explicaciones que no están presentes en los modelos de los autores y, a partir de ellas, decidimos que las explicaciones no son satisfactorias. Sin embargo, esto no es así. En primer lugar, la condición no es tanto de adecuación de las explicaciones sino más bien un rasgo que caracteriza a práctica vinculada con las explicaciones psicológicas en un considerable número de casos; en segundo lugar, no estamos afirmando que las explicaciones sean inadecuadas, sino que no logran hacer justicia a este aspecto.

requerimientos que plantean las explicaciones psicológicas, y hemos concluido que estas doctrinas no son capaces de preservar algunos de los rasgos principales que frecuentemente atribuimos a las explicaciones científicas en psicología. En el siguiente capítulo examinaremos diversos intentos de 'disolver' el problema apelando, en algunos casos, a consideraciones sobre la epistemología de la explicación psicológica.

[99] Existe la posibilidad de que las propiedades mentales sean causalmente eficaces en su propio nivel, aunque incapaces de producir efectos en el mundo físico. Examinaremos en el capítulo VI una propuesta de esta clase.

CAPÍTULO V. ARGUMENTOS DEFLACIONISTAS[100]

1. El problema de la exclusión causal: ¿solución o disolución?

Como se observó en el capítulo II, una estrategia que ha cobrado popularidad en los últimos años frente a la exclusión causal de lo mental es el intento de 'disolver' el problema, en vez de resolverlo; esto es, mostrar que el problema surge de una manera errónea de concebir la forma en que debemos desarrollar nuestro conocimiento de lo mental, y que debería, por lo tanto, ser desechado. Una de las tácticas que se ha empleado con este propósito ha sido mostrar el carácter problemático de los supuestos metafísicos a partir de los cuales el problema se origina, en combinación con la postulación de la necesidad de otorgar prioridad a los aspectos epistémicos de la explicación, dejando de lado los aspectos metafísicos. Este tipo de argumentos es bien conocido, y ha sido respondido en detalle por Kim (1998). Presenta el interés adicional de plantear, de manera más o menos explícita, el rol que la filosofía debe jugar en relación con la investigación fáctica sobre lo mental.

En este capítulo expondremos esquemáticamente, en primer lugar, los lineamientos generales de estas posiciones en las versiones expuestas por Baker (1993), Burge (1993) y Van Gulick (1993), y, en una versión más reciente, por Glymour (1999). En segundo lugar, comentaremos las críticas expuestas por Kim (1998) y Sabatés (2001); sostendremos que las réplicas a estas posiciones son más sólidas que las posiciones mismas, por lo que no pueden constituir una solución (o una 'disolución') al problema. Por último, señalaremos que las insuficiencias en la caracterización de lo que constituye una explicación psicológica aceptable, uno de los pilares de algunos de estos argumentos, constituyen obstáculos difíciles de salvar para esta estrategia.

Sostendremos, en síntesis, que los argumentos deflacionistas, independientemente de la cuestión relativa a la posición frente al problema que resulta adecuada para la filosofía, presentan debilidades que impiden que puedan constituirse en una alternativa promisoria.

---

[100] Expresión utilizada por Kim (1998, p. 59)

2. Variantes de esta estrategia

2.1. Baker, Burge y la prioridad de la explicación por sobre la metafísica

El problema de la causación mental, en opinión de Baker (1993), puede ser planteado a través de la siguiente pregunta: ¿cómo pueden ser causalmente relevantes para la producción de sucesos conductuales las propiedades de contenido [*content-properties*] de los sucesos internos? De la respuesta a esta pregunta se derivará la posibilidad de explicar la conducta en términos de las propiedades de contenido de los estados internos; si tales propiedades son causalmente irrelevantes, y suponiendo ciertas tesis ampliamente admitidas acerca de la naturaleza de la explicación, se sigue que también serán explicativamente irrelevantes. Ahora bien, dado el marco metafísico aceptado por muchos filósofos, ambas preguntas deben ser respondidas de manera desfavorable para el rol causal y explicativo de esas propiedades.

Dos principios, sostiene, constituyen el núcleo del marco metafísico materialista que genera el problema de la causación mental: el principio de clausura causal del mundo físico (CCP) y la noción de superveniencia fuerte (SS):

> (CCP) Toda ejemplificación de una propiedad microfísica que tiene una causa en *t* tiene una causa microfísica completa en *t* (p. 79).
>
> (SS) Necesariamente, para toda ejemplificación de cualquier propiedad *F*, hay una ejemplificación de una propiedad microfísica *G*, y, necesariamente, cualquier cosa que tenga *G* tiene *F* (p. 80).

Ambos principios, observa Baker, no sólo generan un problema para la causación mental, sino que generan también un serio problema para la noción de macrocausación, por lo que resultan afectadas todas las ciencias especiales y aun partes de la física.

El problema de la causación mental se origina cuando notamos que, de cualquier acción, y en virtud de CCP y SS, podemos afirmar que sus determinantes suficientes son propiedades neurofisiológicas, lo cual pone en tela de juicio la pertinencia de afirmaciones como 'hizo *x* en virtud de su deseo de *y*'. Se necesita, entonces, alguna justificación para admitir afirmaciones de esta última clase; en

ausencia de una apropiada, las propiedades mentales parecen ser causalmente irrelevantes.[101]

Con respecto a la macrocausación y a su rol para las ciencias especiales, Baker señala que estos principios implican la idea de que lo que ocurre en el micronivel determina todo lo que ocurre en el macronivel; por lo tanto, si toda ejemplificación de cualquier propiedad sobreviene a partir de la ejemplificación de alguna propiedad microfísica, y la ejemplificación de cualquier propiedad microfísica (que tiene una causa) tiene una causa microfísica completa, se necesita una justificación para la afirmación de que las propiedades macrofísicas son causalmente relevantes para algo. Dado que el marco metafísico establece que la causación es metafísicamente prioritaria sobre la explicación, no puede apelarse al éxito explicativo de las ciencias especiales como forma de obtener esa justificación; el éxito explicativo de tales ciencias es, dados CCP y SS, ilusorio.

La conclusión de Baker es que, dado el marco metafísico que da origen al problema, éste resulta insoluble. Sólo caben dos alternativas ante el problema: o se rechaza una parte de los supuestos metafísicos que configuran el marco a partir del cual surge, o se renuncia a la casi totalidad de las explicaciones que pueden ofrecerse para cualquier suceso, incluyendo las del sentido común y las de las ciencias especiales. Consecuencias tan devastadoras para la ciencia y el sentido común, considera, las convierten en una verdadera *reductio ad absurdum* del marco metafísico.

Si bien Baker afirma que SS constituye una especulación hueca, un supuesto metafísico cuyo propósito es principalmente cubrir la necesidad de una tesis totalizante, su rechazo se dirige principalmente al principio CCP y a la noción de causalidad postulada por Kim, a la cual califica de 'inaceptablemente estrecha'. Su propuesta consiste en tomar como punto de partida filosófico no una doctrina metafísica, sino la explicación:

[101] Baker analiza, y desecha, un posible principio para lograr este objetivo: 'P Si una ejemplificación de la propiedad *G* en *t* es una causa completa de una ejemplificación de *H* en *t'* (*t'* > *t*), y, necesariamente, cualquier ejemplificación de *G* en *t* es suficiente para una ejemplificación de *F* en *t*, entonces la ejemplificación de *F* es causalmente relevante para la ejemplificación de *H* en *t"* (p. 84). Pero este principio admite contraejemplos: que una mujer tenga sarampión durante el embarazo es una causa suficiente para la ejemplificación de defectos de nacimiento, y, necesariamente, cualquiera que tenga sarampión en *t* tiene manchas rojas en *t*. Se seguiría de aplicar el principio P que el tener manchas rojas es causalmente relevante para que el niño tenga defectos de nacimiento, lo cual es inaceptable.

> ¿Cómo, entonces, entendemos la causación? Mi sugerencia es tomar como punto de partida filosófico no una doctrina acerca de la naturaleza de la causación o la realidad, sino un rango de explicaciones que hemos hallado dignas de aceptación. Este incluye, predominante pero no exclusivamente, explicaciones científicas. También incluye las explicaciones comunes que encontramos en la vida cotidiana (...) Construyendo explicaciones como respuestas a preguntas de 'por qué', quizás con algunas restricciones acerca de lo que se considera una respuesta adecuada, mi propuesta es comenzar con explicaciones que ganen su sustento [*earn their keep*], más que con la metafísica, la cual me parece un supuesto gratuito [*freeloader*] que sólo interfiere con el trabajo real (pp. 92-93).

La causación se convierte, en la propuesta de Baker, en un concepto explicativo, adoptando una posición fuertemente opuesta a la concepción metafísica que subordina la explicación a la causación. La opción elegida es, entonces, no resolver el problema de la causación mental sino disolverlo por medio del rechazo del marco metafísico a partir del cual se origina.

Baker observa, finalmente, que si bien su propuesta tiene un marcado tono pragmático, no implica una posición antirrealista; no pretende igualar lo que es real con lo que se requiere para formular predicciones y explicaciones. Lo que sí es correcto, a su modo de ver, es que no hay mejor acceso a la realidad que el que se requiere para el éxito cognitivo, interpretado de manera tal que incluya lo que sea cognitivamente necesario para el logro de metas tanto en la ciencia cuanto en la vida cotidiana.

La línea argumentativa seguida por Burge (1993) es básicamente similar a la de Baker. En su opinión, las preocupaciones relacionadas con el presunto epifenomenismo de lo mental son 'sintomáticas de un conjunto erróneo de prioridades filosóficas' (p. 97). Burge considera que se ha concedido muy poco peso a la reflexión acerca de la práctica explicativa, la cual constituye el punto de partida más plausible para la reflexión mente-cuerpo, y demasiado a la metafísica materialista.

El argumento general de Burge puede ser dividido en tres partes. En la primera presenta dos argumentos que conducen aparentemente al epifenomenismo, basados en premisas comúnmente aceptadas por los materialistas, y señala ciertas debilidades en ellos. En particular, dos de las premisas (una por cada argumento) son especialmente cuestionadas: la noción de 'poder causal', en el primer argumento,

y la concepción de los sistemas físicos como causalmente cerrados. En su opinión, la causación mental y los poderes causales de lo mental no pueden ser comprendidos concentrándose en las propiedades caracterizadas por las ciencias físicas; por el contrario, nuestro entendimiento de la causación mental deriva primordialmente de nuestra comprensión de las explicaciones mentales, independientemente de nuestro conocimiento de los procesos subyacentes. Los enfoques materialistas, concluye, han ensanchado demasiado la brecha existente entre la metafísica de la causación mental y lo que realmente sabemos sobre la naturaleza de la causación mental, conocimiento que deriva casi por completo de nuestras observaciones y explicaciones mentalistas.

En segundo lugar, Burge examina las bases de la metafísica materialista que constituyen el fundamento de las preocupaciones acerca del epifenomenismo. Si bien admite que existen buenas razones para considerar que existen estados y procesos físicos, químicos, biológicos y neurales subyacentes a nuestros estados y procesos mentales, y que esos estados y procesos obedecen a sus propias leyes, la naturaleza de la relación entre los sucesos mentales y los físicos dista de estar clara. Presenta un argumento tendiente a mostrar la inadecuación de la tesis materialista de la identidad de casos, que constituye el intento más ampliamente aceptado de clarificar la relación mencionada. Este argumento se basa en los conocidos experimentos mentales desarrollados por Putnam y por el propio Burge, a partir de los cuales hay una fuerte presunción en favor de la idea de que es posible para dos personas tener pensamientos con diferentes contenidos aun cuando todos los sucesos que ocurran en los organismos individuales que sean candidatos plausibles para la identificación con esos sucesos mentales sean los mismos. Dado que encuentra a las premisas de su argumento fuertemente plausibles, y libres de contraargumentos, concluye que la base metafísica más común para las preocupaciones epifenomenistas es muy débil.

Por último, Burge examina las demandas usuales de los materialistas de que toda explicación de sucesos físicos debe proveer de un mecanismo causal por el cual el suceso acontece (o requisitos similares). Siguiendo este modelo, parecería que la explicación de la causación mental debe de alguna manera interferir o alterar los procesos puramente físicos. Pero debido a que los antecedentes físicos de un suceso

parecen ser suficientes para sus efectos, y a que esta intervención de causas mentales nos haría dudar de la adecuación de las formas actuales de explicación física, tal posibilidad debería ser rechazada. Burge observa que este argumento tiene cierta fuerza, pero no toda la que usualmente se le atribuye. Pensar que la causalidad mental debe interferir en alguna forma con la causalidad física depende de concebir la primera bajo un modelo físico; una causa, cualquiera que fuere, debería transmitir un fragmento de energía o ejercer una fuerza sobre el efecto. Pero la adecuación de este modelo para concebir la causalidad mental, señala, es justamente lo que está en discusión. Tenemos razones para pensar que las explicaciones mentalistas y fisicalistas no tienen por qué interferir mutuamente. Las explicaciones mentalistas funcionan y no entran en conflicto con las explicaciones fisicalistas.

Si bien deben existir relaciones sistemáticas entre los sucesos mentales y los procesos físicos subyacentes (Burge se inclina por alguna clase de relación amplia de superveniencia: ningún cambio en un estado mental puede tener lugar sin un cambio en los estados físicos), hacer de esto una inferencia hacia el materialismo es una especulación metafísica que ha devenido erróneamente en una especie de lugar común.

La base para nuestra comprensión de la causación mental no es, entonces, la metafísica materialista, sino nuestra creencia en la adecuación de nuestras explicaciones causales mentalistas:

> Ya que la explicación mentalista produce conocimiento y comprensión, y ya que esa explicación es (a veces) causal, podemos creer firmemente que la causación mente-cuerpo es una parte del mundo. La manera primaria de entender tal causación es comprendiendo las afirmaciones causales y explicativas mentalistas en el sentido ordinario, no filosófico, de 'comprensión'. Cuánto más esclarecimiento puedan ofrecer la filosofía o la neuropsicología es algo que está por verse. En cualquier caso, la explicación mentalista y la causación mental no necesitan validación desde la metafísica materialista (p. 117).

> Pienso que es más natural y fructífero comenzar por suponer, refutable [*defeasible*] pero firmemente, que las atribuciones de estados mentales intencionales son centrales en las explicaciones psicológicas en la vida cotidiana y en varias partes de la psicología. Podemos suponer también que los sucesos mentales son a menudo causas y que la explicación psicológica es a menudo una forma de explicación causal. Dados estos supuestos, la 'preocupación' relativa al epifenomenismo parece muy remota (p. 118).

El epifenomenismo, concluye Burge, no constituye una opción metafísica que deba ser tomada seriamente en cuenta; sería mejor considerarlo, al igual que al escepticismo, como un instrumento útil para clarificar nuestras creencias más profundas.

2.2. El socavamiento de la primacía de lo físico: la propuesta de Van Gulick

Como otros autores, Van Gulick (1993), comienza su propuesta advirtiendo acerca de la reciente difusión de preocupaciones epifenomenistas entre los filósofos dedicados al estudio de lo mental. Pese a que el epifenomenismo de las propiedades mentales es un 'híbrido antiintuitivo y no muy atractivo', observa, existe una serie de argumentos que parecen conducirnos a él. Tres argumentos son ofrecidos como muestra de esta tendencia: el argumento de las leyes estrictas, el argumento de la exclusión y el argumento de la inexistencia de la superveniencia individual; una extensión de este último intenta mostrar que las teorías psicológicas no deberían individuar estados en términos de sus propiedades de contenido amplio.

En opinión de Van Gulick, el argumento de la exclusión está mucho más cerca del núcleo del problema que los restantes; a diferencia del argumento de las leyes estrictas, no requiere de ninguna suposición relativa a la manera en que dos sucesos están causalmente conectados. Por otra parte, no ofrece flancos débiles manifiestos, ya que descansa en ciertas premisas ampliamente aceptadas dentro de un marco fisicalista. Estas premisas son el principio de que todo suceso es un suceso físico y el principio de clausura causal del mundo físico. En vez de atacar directamente alguno de los supuestos del argumento, la estrategia que Van Gulick elige para debilitarlo consiste en profundizar la cuestión de la relevancia causal, con el fin de minar sus supuestos ocultos.

Una manera de responder el argumento de la exclusión, considera Van Gulick, es poner de manifiesto el rol de las leyes y de las explicaciones causales en nuestra economía cognitiva. El rol de las leyes causales es el de proveer los medios para formular proyecciones hacia el futuro y formar planes de acción. Si bien las leyes causales ideales para nuestros propósitos explicativos son leyes simples, confiables y precisas, leyes de esta clase raramente están disponibles; en las explicaciones causales, por lo tanto, la precisión debe a menudo ser sacrificada

frente a las limitaciones pragmáticas acerca de lo que podemos detectar o comprender. Las explicaciones intencionales podrán ser menos precisas y confiables que las explicaciones microfísicas de los mismos hechos; sin embargo, están disponibles en la práctica y relacionan propiedades que podemos detectar y que son relevantes para nuestros intereses. El defensor de la exclusión podrá replicar, observa Van Gulick, que esta respuesta no resuelve la cuestión de si las propiedades mentales son causalmente potentes; permanece sin respuesta el problema de que, en las explicaciones mentales, el trabajo causal real es realizado por las propiedades físicas subyacentes. Esta réplica del defensor de la exclusión lleva, en última instancia, a sostener la impotencia causal no sólo de las propiedades mentales, sino también de las propiedades de todas las 'ciencias especiales'.[102] Si este fuera el caso, una respuesta posible consistiría en preguntarse por qué las propiedades mentales deberían gozar de un standard privilegiado con respecto a las propiedades de otras ciencias especiales. Pero esta respuesta, considera, no es suficiente para quienes estén preocupados por el epifenomenismo de lo mental.

La estrategia que Van Gulick adopta no es mostrar que las propiedades de las ciencias especiales tienen el status de las propiedades físicas, sino mostrar que las propiedades físicas *no tienen* el status especial que usualmente se les atribuye. Van Gulick intenta demostrar que las leyes y explicaciones de las ciencias especiales se refieren a patrones de organización de las propiedades de nivel superior que gozan de relativa independencia con respecto a la organización de las propiedades físicas subyacentes. Estos patrones no son menos reales y causalmente potentes que las propiedades y leyes físicas a partir de las cuales se organizan. Ciertas características que poseen (entre otras, son rasgos recurrentes y estables del mundo; muchos de ellos son estables pese a las variaciones en sus propiedades físicas subyacentes; pueden afectar los poderes causales de sus constituyentes por medio de la activación selectiva) hacen que estos patrones puedan tener un grado de independencia de sus propiedades constituyentes y afectarlas, no mediante alteración de las leyes físicas subyacentes, sino mediante su activación selectiva.

[102] A una conclusión similar llega Baker con respecto a la eficacia causal de las macropropiedades. *Cfr*. su (1993), pp. 86-90.

Por otra parte, Van Gulick señala que las propiedades físicas que se encuentran en leyes estrictas, en última instancia, no son más que 'estados estables, recurrentes y auto-mantenedores del flujo cuántico de una realidad irreductiblemente probabilística y estadística' (p. 254). Se organizan en patrones estables cuyas interrelaciones se aproximan en un grado muy alto a regularidades deterministas. Su rigor determinista, observa, produce en nosotros la ilusión de que estamos tratando con algo más que una representación de algunos aspectos de la organización del espacio-tiempo, algo más tangible, real y objetivo. Pero esto no implica, concluye, que tales propiedades deban jugar un rol único (ni siquiera especial) de tipo metafísico en determinar la organización temporal del mundo.

A partir de estos argumentos, Van Gulick considera que el argumento de la exclusión se encuentra suficientemente debilitado; el pasaje de las premisas a la conclusión relativa a los poderes causales de las propiedades mentales no está garantizado, ya que no se han considerado apropiadamente los roles causales desempeñados por las propiedades de orden superior.

2.3. Glymour: la estrategia 'humpty dumpty'

La posición de Glymour (1999) representa una versión más reciente de la estrategia deflacionista; coincide con las anteriores en que la manera metafísica tradicional de comprender la causación mental constituye un error básico. El problema de la causación mental planteado por Kim y por otros autores, observa, representa un acabado ejemplo de lo que denomina 'corriente metafísica principal en filosofía de la mente' [*mainstream-metaphysical-philosophy-of-mind*]. En esta corriente los proyectos filosóficos se encuentran 'amurallados' en relación con cualquier uso real de la matemática o la ciencia.

Creemos que es verdad, sostiene Glymour, que nuestros pensamientos, creencias, deseos, planes, esperanzas y temores causan nuestras acciones, o al menos causan algo de lo que hacemos. El desafío de la causación mental es mostrar cómo es razonable creer en ella. Ante las diversas estrategias que, en su opinión, pueden adoptarse ante este desafío, Glymour opta por la que denomina 'estrategia Humpty Dumpty': si bien el argumento en contra de la causación mental es sólido, hemos estado hablando por un largo tiempo de manera muy satisfactoria sobre los

pensamientos como si fuesen causas, y planeamos continuar haciéndolo; la ciencia no tiene nada que ver con él.

Debe notarse, observa Glymour, que en sí misma la revelación científica de los mecanismos físicos que median la creencia y el deseo no socavan, y no deberían hacerlo, la convicción de que los estados mentales causan las acciones. No pensamos, sostiene, que la articulación de un mecanismo de causación socave una causa. Por ejemplo, no concluimos que un ácido no causa que el bicarbonato de sodio libere un gas porque conozcamos las propiedades moleculares que los ácidos comparten, la composición molecular del bicarbonato de sodio y las reacciones químicas que ocurren cuando el ácido y el bicarbonato son combinados.

Glymour expresa con claridad su punto de vista con respecto a nuestra comprensión de la causalidad:

> Nuestra comprensión de la causación deriva de nuestra experiencia con la producción de efectos por la manipulación de causas, ya sea por nosotros mismos, por otros, o por las circunstancias, y a partir de la extrapolación de relaciones de dependencia en los sistemas que no podemos manipular y aun en aquellos que la Naturaleza no manipula. Cuando pido a mi hija obediente (...) que cierre la puerta, causo en ella la creencia de que la puerta está abierta y el deseo de cerrarla, los cuales causan conjuntamente que ella cierre la puerta (...) Nada es más natural, o más científico, que suponer que la creencia que causamos causa la acción que la sigue (1999, p. 470).

Glymour observa, no obstante, que la evidencia disponible subdetermina el carácter del proceso. Podría ocurrir que el pedido causara tanto la creencia de la hija y su acción, pero no que la creencia causara la acción. La convicción de que los estados mentales concientes causan las acciones estaría subdeterminada por la evidencia de que los mecanismos físicos causan los estados mentales concientes y los mecanismos físicos causan la acción, y estos mecanismos son separables, de manera que con intervenciones apropiadas las acciones podrían ser causadas sin la ocurrencia de los estados mentales. Glymour considera que se está acumulando evidencia, si bien aún insuficiente, de que los mecanismos que producen los estados mentales concientes podrían ser de hecho separables de aquellos que producen las acciones que acostumbramos atribuir a pensamientos concientes. Por ejemplo, existe evidencia de que los procesos fisiológicos a partir de los cuales la acción

voluntaria puede ser predicha preceden cualquier experiencia de volición o decisión por 300 milisegundos o más.[103]

Pero la determinación de si es posible que los mecanismos que subyacen a los estados mentales concientes sean separables de los mecanismos de las acciones intencionales es una cuestión empírica, y no metafísica. Todo el análisis de Kim, finaliza Glymour, es un ajuste del interior de la casa constituida por la 'corriente metafísica principal en filosofía de la mente', reacomodando algunas piezas, descartando otras y adicionando algunas nuevas, mientras los fundamentos están deteriorados. Es mejor, concluye, construir en la ciencia.

3. La insatisfactoriedad de los argumentos deflacionistas: algunas réplicas

De la descripción precedente puede apreciarse que las estrategias deflacionistas coinciden en algunos aspectos centrales y difieren en otros. Con respecto a los primeros, posiblemente la coincidencia más saliente es en el diagnóstico: la metafísica materialista no puede constituir la piedra basal para la reflexión acerca de la causación mental. Si los problemas de la causación mental tienen una solución, ésta no provendrá de las doctrinas que asignan una posición de privilegio al mundo físico. Sin embargo, difieren en cierta medida en la estrategia a adoptar. Mientras que Baker y Burge coinciden en la necesidad de partir de la práctica explicativa, tanto científica cuanto de sentido común, Van Gulick hace del rechazo al supuesto privilegio metafísico de las propiedades y leyes físicas su argumento principal; por último, Glymour parece situar todo el peso del problema en la práctica científica, simplemente retirándolo del campo de la filosofía.

Debido a estas diferencias la clase de objeciones que deben enfrentar difiere considerablemente. Es de destacar, no obstante, que los argumentos deflacionistas a menudo resultan insuficientes aun para autores que rechazan que el problema de la exclusión causal sea insoluble dentro del marco del materialismo no reduccionista.[104] Parece haber varias razones por las cuales esta opción no es satisfactoria.

En los apartados que siguen nos referiremos principalmente a las posiciones de Baker y Burge, ya que son quienes con mayor énfasis han apelado a la necesidad

[103] Si bien Glymour no lo menciona, es posible que esté haciendo referencia a las experiencias de Libet. *Cfr*. al respecto el capítulo II.

de recurrir a la práctica explicativa como clave para resolver el problema. En primer lugar, describiremos la crítica expuesta por Sabatés (2001), quien considera que la estrategia empleada por estos autores desemboca en una suerte de irrealismo acerca de lo mental; respecto de esto, señalaremos que parece plausible pensar que esta posición es inconsistente con el éxito explicativo de la psicología que esta posición requiere. En segundo lugar, describiremos la réplica de Kim a esta estrategia, a la que consideraremos sólida. Por último, retomaremos el problema que afrontan Baker y Burge de proporcionar criterios de éxito explicativo que les permitan recurrir con posibilidades de éxito a la psicología para fundar su propuesta.

3.1. ¿Realismo o instrumentalismo?

Sabatés (2001) sugiere que una vez que se invierte la prioridad causación/explicación[105] el resultado final será plausiblemente un enfoque irrealista acerca de la mente, quizás de tipo retentivo. Pero muchos realistas sostendrían, por el contrario, que dentro de un marco realista la posibilidad de la causación mental es una precondición de la posibilidad de lograr explicaciones psicológicas *causales*. Sabatés sugiere que, dado el enfoque de la 'primacía explicativa' sostenido por Baker y Burge, los predicados mentales podrían ser meras herramientas explicativas y predictivas, pero la realidad de lo mental no parecería estar entre las tesis a ser sostenidas. La realidad de lo mental podría ser defendida proponiendo una tesis del siguiente tipo: 'una propiedad es real si y sólo si figura en explicaciones exitosas'. Pero, observa, puede suponerse plausiblemente que no es éste el tipo de concepción sobre la realidad de las propiedades que un realista acerca de lo mental quisiera defender. Esta estrategia supuestamente compatibilista, en su opinión, resulta ser

[104] *Cfr*. Marras (2000).

[105] Kim (1987), al formular su distinción entre realismo e irrealismo explicativo, adjudica a éste último la característica de dar prioridad a la explicación por sobre la causación, y atribuye una posición tal a Hanson: 'Para Hanson, las relaciones causales dependen esencialmente de un entrelazamiento conceptual apropiado de nuestras descripciones tal como son provistas por las teorías que aceptamos (...) [C]oncibe la relación causal entre *x* e *y* como derivativa de una relación inferencial de *x* a *y*, y la relación inferencial está íntimamente asociada con la explicación; es también evidente que no toma la dependencia de la causación de la inferencia y la explicación como meramente epistemológica. Si uno acepta esta concepción de la causación y la explicación causal, no hay nada realista en la posición de que las explicaciones causales sólo tienen lugar en caso de que la relación causal tenga lugar. Las relaciones causales, bajo este enfoque, dependen de conexiones inferencial-explicativas que son primarias y más básicas' (pp. 227-228).

una postura incompatibilista, en la cual la tesis a rechazar es el realismo acerca de lo mental.[106]

Si la interpretación de Sabatés es correcta, podría pensarse entonces que la estrategia de la prioridad explicativa desemboca en una variante del instrumentalismo semántico restringido al ámbito de lo mental. Los predicados mentales sobrevivirían por razones pragmáticas; permitirían la predicción y la explicación, pero no harían referencia a propiedades reales.[107] Pero esta posibilidad es inconsistente. Por el contrario, parecería que Baker y Burge se ven en la necesidad de defender el criterio de realidad basado en el éxito explicativo. Veremos enseguida que este criterio es sin duda problemático, pero por ahora fijemos una condición mínima que parece necesaria para poder afirmar que una explicación es exitosa: una explicación exitosa implica que las propiedades a las que se hace referencia en el *explanans* son reales. Hemos señalado en el capítulo anterior que negar esta condición impediría, entre otras consecuencias indeseables, rechazar explicaciones basadas en el éter o el flogisto.[108] Pero si esta condición mínima se acepta, sería inconsistente sostener luego que los predicados mentales no hacen referencia a propiedades reales. Su posición, entonces, no puede ser considerada como una clase de instrumentalismo semántico. Resulta plausible, entonces, presentarla como una clase de realismo con un criterio *sui generis* de realidad. Este criterio, como veremos luego, es extremadamente controvertible.

3.2. El carácter inevitable de la metafísica: la respuesta de Kim

En opinión de Kim, ninguna de las estrategias 'deflacionistas' (entre las cuales incluye a las de Baker y Burge) tiene realmente éxito; el problema de la

[106] Las afirmaciones de Baker relativas a que no intenta igualar lo que existe con lo que resulte necesario para la explicación y la predicción resultan un tanto desconcertantes. Si, como ella sostiene, nuestro mejor (o único) acceso a la comprensión de la naturaleza de lo mental es lo que se requiere para el éxito cognitivo, parecería ser consistente con esta propuesta la admisión de que lo real *es* lo que se requiere para el éxito cognitivo.

[107] Parece dudoso, sin embargo, que el instrumentalismo semántico permita más que la predicción. En efecto, si las entidades teóricas no son más que ficciones útiles para la producción de predicciones, difícilmente pueda afirmarse que, sobre la base de tales entidades ficticias, tengamos una explicación satisfactoria de algún fenómeno.

[108] Esto no quiere decir que las explicaciones basadas en el éter o el flogisto no puedan resultar insatisfactorias por otras razones; sin embargo, ninguna razón parece mejor para rechazar una explicación que el conocimiento de que las entidades o propiedades explicativas a las cuales se hace referencia en el *explanans* no existen (salvo, quizás, que el *explanandum* mismo sea falso).

causación mental no será resuelto simplemente efectuando ciertas reparaciones poco costosas en algunos lugares.

Kim considera que los planteos de Burge y Baker son correctos en un aspecto: es verdad que 'nuestra confianza en la verdad de las explicaciones intencionales familiares excede nuestro compromiso con cualquier principio metafísico recóndito' (1998, p. 61). Sin embargo, advierte, nuestras preocupaciones no son evidenciales o epistemológicas. El problema de la causación mental es, en primer lugar, un problema metafísico: es el problema de mostrar *cómo* la causación mental es posible, no *si* es posible. Al formular la pregunta de *cómo*, estamos suponiendo que la pregunta de *si* ya ha sido respondida afirmativamente. La cuestión es, observa, cómo hacer que nuestra metafísica sea consistente con la causación mental, y la elección que debe realizarse es entre diversas alternativas metafísicas, no entre algunos recónditos principios metafísicos por una parte y algunas prácticas o principios epistemológicos apreciados por la otra. Esto no implica decir que la metafísica y la epistemología son necesariamente independientes; por el contrario, ambas están relacionadas en esta área, y las elecciones que hacemos en una pueden requerir ajustes en la restante.

Cuando Burge afirma que nuestra comprensión de la causación mental proviene de nuestra comprensión de la explicación mentalista, señala Kim, está sosteniendo que las explicaciones mentalistas son a menudo explicaciones causales. Sin embargo, si se recuerdan los prolongados debates en las décadas del '50 y el '60 respecto de si las explicaciones de la acción basadas en el esquema creencia-deseo son causales, se observará que la posición según la cual tales explicaciones no son causales fue dominante la mayor parte de ese tiempo, hasta que la teoría causalista debida a Davidson se transformó en la nueva ortodoxia. Kim señala agudamente que el desacuerdo no tenía lugar con respecto a la práctica explicativa, sino sobre su naturaleza y razón fundamental [*rationale*]. Respecto de esto, advierte, estos debates son suficientes para mostrar que no puede aislarse fácilmente la práctica explicativa de la comprensión racional de la conducta de complicaciones metafísicas. La pregunta acerca de si las explicaciones racionalizadoras son una clase de explicación causal involucra temas metafísicos sustantivos.

Aun cuando se haya respondido a la cuestión de manera favorable a la posición causalista, la metafísica no desaparecerá. La idea de explicación causal, prosigue Kim, presupone la idea de que el suceso al cual se hace referencia en la explicación causal es una causa del fenómeno que debe ser explicado. En otros términos, si *c* (o una descripción o representación de *c*) explica causalmente *e*, entonces *c* debe ser una causa de *e*, lo cual parece incontrovertible.

En presuntos casos de causación de lo mental a lo físico, los problemas metafísicos emergen rápidamente en varios puntos. Para cualquier caso en el cual consideramos que una conducta ha sido causada por un deseo (causación de lo mental a lo físico), tendremos una concepción alternativa que invoca una cadena de sucesos neurofisiológicos como causa de esa conducta. ¿Cuál es la relación entre ambas explicaciones? En este punto, Kim recurre al esquema a través del cual describe la competencia de explicaciones causales expuesto en su (1989).[109] Sólo puede tratarse de: a) un caso de sobredeterminación; b) un caso de causas parciales; c) un caso en el cual una causa sea parte de la otra; d) una sola causa bajo diferentes descripciones; e) una causa (presumiblemente la causa mental) se reduce a la otra; y f) el status causal de una de las causas (presumiblemente la mental) es, en algún sentido, dependiente de la otra causa (neural). La presencia de dos o más explicaciones causales crea una situación de inestabilidad que sólo puede disiparse determinando cómo esas explicaciones se relacionan entre sí.

Burge podría replicar que las explicaciones intencionales y las explicaciones neurofisiológicas no necesitan competir entre sí, y que de hecho no lo hacen. Pero lo que no toma en consideración, observa Kim, es que 'dos o más explicaciones pueden ser explicaciones rivales aun cuando sus premisas explicativas sean mutuamente consistentes y de hecho todas verdaderas, en tanto se propongan explicar (en particular explicar causalmente) un único *explanandum*' (p. 65). Que esas explicaciones surjan en distintas áreas de investigación, que estén en diferentes niveles de análisis o descripción, o que sean respuestas a distintas preocupaciones epistémicas o pragmáticas, no hace ninguna diferencia.

La metafísica, sostiene Kim, es el dominio en el cual los diferentes lenguajes, teorías, explicaciones y sistemas conceptuales se unen, y en el que sus relaciones

ontológicas mutuas son ordenadas y clarificadas. Pero la real dimensión del problema puede ser apreciada, en su opinión, sin necesidad de aceptar demasiadas tesis metafísicas:

> El problema de la exclusión causal/explicativa surge si hay casos de explicaciones psicológicas de conducta física acerca de los cuales estamos preparados para creer que el efecto físico también tiene, o debe tener, una explicación causal física. Y no necesitamos suscribir una doctrina general de la clausura causal del dominio físico para creer que de hecho debe haber tales casos; la clausura causal de lo físico sólo generaliza el problema de la exclusión a todas las explicaciones mentalistas de conductas físicas (...) Surge de la noción misma de explicación causal y de lo que me impacta como una comprensión perfectamente intuitiva y común de la relación causal (1998, pp. 66-67).

Por último, Kim observa que tanto Burge como Baker admiten alguna forma de dependencia o superveniencia mente-cuerpo. Debido a esto, enfrentan la necesidad de responder al argumento de la superveniencia: deberían mostrar cómo resistir a este razonamiento que, aparentemente, conduce al epifenomenismo acerca de lo mental.

Estas réplicas de Kim a las objeciones de Baker y Burge nos parecen sólidas. Sin embargo, existe otro punto débil en las posiciones de estos autores, que analizaremos a continuación.

4. El éxito explicativo

Hemos señalado en el capítulo IV algunas dificultades vinculadas con la caracterización de la noción de 'éxito explicativo'. Dado que Baker y Burge recurren explícitamente a las explicaciones psicológicas como base para la comprensión de la causalidad mental, será conveniente profundizar en este concepto. Hemos visto que Baker sugiere comprender la causación mental a partir de 'un rango de explicaciones que hemos hallado dignas de aceptación'; principalmente explicaciones científicas. Burge, por su parte, considera que partir de la práctica explicativa es la manera adecuada de razonar acerca de la causación mental.

Ahora bien, Baker parece advertir las dificultades derivadas de aceptar incondicionalmente cualquier explicación mentalista como base para la comprensión

[109] *Supra*, p. 23.

de la causalidad mental, y sugiere, como se recordará, establecer algunas restricciones acerca de lo que se considera una explicación aceptable. Burge no propone establecer una restricción similar, aunque es plausible suponer que, interrogado sobre la cuestión, estaría de acuerdo en que no toda explicación psicológica puede cumplir la función que pretende asignarle. Una razón trivial para este acuerdo sería sencillamente porque no parece factible que todas las explicaciones que encontramos en la psicología sean causales; una razón más profunda sería que parece muy plausible suponer que no todas las explicaciones son aceptables. Supondremos, entonces, que la necesidad de limitar el rango de explicaciones que se consideren aceptables es común a las dos propuestas. Esta cualificación aparentemente inocente encierra un problema que echa por tierra mucho del posible atractivo que pudiera tener su propuesta. Dos interrogantes surgen inmediatamente: en primer lugar, ¿cuáles son las restricciones que debemos introducir para considerar que una explicación es adecuada? En segundo lugar y quizás más importante para el tema en discusión, ¿podemos responder a la pregunta anterior sin reflexionar sobre la metafísica de la explicación? Hemos considerado que los argumentos de Kim son sólidos en fundamentar una respuesta negativa a esta última cuestión. Sin embargo supongamos, *for the sake of the argument*, que podemos reflexionar acerca de las restricciones que impondremos a las explicaciones sin hacer metafísica.

El suponer que podemos seleccionar, dentro del conjunto de las explicaciones provistas por la psicología, un rango de ellas que consideramos 'dignas de aceptación' es suponer (razonablemente) que esta disciplina ha tenido un cierto éxito explicativo. Esto es, ha logrado explicar satisfactoriamente (causalmente, para los intereses de Baker y Burge) un espectro suficientemente amplio de hechos. No sólo debemos suponer que *ciertas teorías* han tenido éxito explicativo; debido a que la psicología se ha caracterizado por la coexistencia de diversas teorías sobre muchos fenómenos, y no hay una teoría unificada que provea explicaciones causales de conjuntos muy amplios de hechos, es razonable suponer que las explicaciones causales satisfactorias deberán extraerse de diversas teorías. Ahora bien, ¿qué significa tener éxito explicativo? La noción de éxito explicativo de una disciplina (o aun de una teoría) es compleja inclusive en los casos de disciplinas más desarrolladas

que la psicología.[110] Son conocidos los casos de teorías físicas que presentan méritos indudables, por ejemplo, predictivos y tecnológicos, pero resulta dudoso que sus explicaciones sean satisfactorias en la misma medida. El caso de la mecánica cuántica es ilustrativo a este respecto. Por otra parte, el éxito explicativo, a diferencia del éxito predictivo y el éxito práctico, no parece ser inmune a los cambios teóricos. Nadie consideraría hoy en día satisfactorias a las explicaciones de fenómenos físicos proporcionadas a partir de la mecánica newtoniana. Sin embargo, para propósitos prácticos, podemos seguir utilizándola para la formulación de predicciones.

Hemos señalado en el capítulo anterior que podría afirmarse que la psicología, considerada globalmente, tiene éxito explicativo, si se considera la cantidad y diversidad de fenómenos que son explicados dentro de su marco, fenómenos que carecen de una explicación dentro de otras disciplinas científicas y dentro de la psicología de sentido común. El éxito explicativo consistiría, entonces, en proporcionar explicaciones de fenómenos que hasta el momento carecen de explicación. Pero tal vez sea ésta una noción demasiado liberal, y posiblemente inapropiada, de éxito explicativo. El éxito explicativo no puede consistir en una cuestión meramente cuantitativa (esto es, únicamente en términos de la cantidad de explicaciones que se posean acerca de los fenómenos), que no considere las características de tales explicaciones.

Es posible que un criterio tal sea susceptible de críticas análogas a las realizadas a la propuesta de Laudan (1977) de considerar como problemas científicos legítimos a aquella clase de problemas que no describen estados de cosas reales (por ejemplo, explicar las propiedades y el comportamiento de las serpientes marinas, o explicar cómo la sangre caliente de la cabra tenía la propiedad de partir diamantes):

[110] No carece de interés señalar que, en algunas caracterizaciones de la empresa científica, el éxito explicativo no sea un criterio para decidir si una teoría es o no aceptable. Newton-Smith (1981) señala que si bien el objetivo de la ciencia es producir verdades explicativas, 'tendemos a prestar más atención a la generación con éxito de nuevas predicciones corroboradas que a la explicación de hechos conocidos, porque, dado un conjunto finito de hechos conocidos, podríamos, con ingenio, concebir alguna teoría – que podría ser muy compleja y embarazosa- a partir de la cual se pudieran derivar esos hechos. Nuestra primera protección contra tales teorías *ad hoc* consiste en el requisito de que aparezcan nuevas predicciones corroboradas' (p. 243).

> A menos que la verdad desempeñe un papel regulador, cada uno de nosotros puede escoger, a su antojo, su propio conjunto de enunciados, que para nosotros son enunciados de problemas porque así hemos decidido considerarlos. Luego, cada uno de nosotros erige sus propias teorías para resolver esos problemas (...) Nos encontraríamos con el poco edificante espectáculo de una pluralidad de conjuntos de problemas en flotación libre y sus teorías asociadas, a algunas de las cuales les corresponderían idénticos valores en la escala para la evaluación de teorías (...) Este modelo convierte a toda la empresa científica en un absurdo (Newton-Smith, 1981, pp. 208-09).

También parece pertinente la observación de que, según esta metodología

> [U]na teoría que resuelva muchos seudoproblemas puede ser mejor que una que resuelva problemas reales, pero en menor cantidad; las teorías que resulten ser las mejores pueden proporcionarnos menos conocimiento del mundo que las peores. Siendo esto así no se entiende, entre otras cosas, por qué dice que en la historia de la ciencia ha habido progreso *cognoscitivo* (Comesaña, 1995, p. 27. Cursivas del autor).

Suponer que una ciencia (para este caso, la psicología) tiene éxito explicativo solamente porque proporciona explicaciones a muchos problemas, sin examinar la legitimidad de tales problemas (y de las explicaciones correspondientes), parece conducir a las mismas consecuencias indeseables. Por ejemplo, aplicando este criterio resultaría difícil negar que el psicoanálisis, pese a las conocidas objeciones teórico-metodológicas que se han presentado en su contra, tiene mayor éxito explicativo que la mayoría (si no todas) las teorías rivales en su campo. La situación puede ser peor aún si se consideran las explicaciones de fenómenos teóricos. Una teoría compleja, que postule muchas entidades teóricas y presuntas relaciones objetivas entre tales entidades puede aventajar claramente a sus rivales en el número de explicaciones que proporciona, simplemente por el hecho de que propone más entidades y relaciones que sus rivales, aun cuando muchas de estas entidades puedan resultar, a la postre, ficticias.[111] La teoría más explicativa podría revelarse, a la postre, como la más insatisfactoria. Parece claro que una noción plausible de éxito explicativo debe descartar esta posibilidad.

Una primera restricción obvia para determinar un criterio de éxito explicativo es, entonces, que los *explananda* hagan referencia a estados de cosas reales. Sin

[111] La postulación, por parte del psicoanálisis, de los instintos de vida y de muerte y de sus interacciones como principios explicativos fundamentales de una vasta serie de fenómenos podría ser un ejemplo apropiado de esta posibilidad.

embargo, aquí entra en juego una segunda dificultad, que es la competencia explicativa. El hecho de que existan diferentes entidades y propiedades explicativas no se debe solamente a la amplitud del rango de fenómenos a explicar, sino también a la coexistencia de teorías explicativas sobre esos mismos *explananda*, que proponen entidades, propiedades y leyes (en caso de que las haya) diferentes como base para la explicación. Tales teorías, además, no compartirán los mismos criterios acerca de lo que constituye una explicación adecuada. Baker y Burge parecen encontrarse con una tarea ímproba por delante: deberían proponer criterios temáticamente neutrales acerca de lo que constituye una explicación psicológica causal adecuada, y sin que estos criterios incluyan consideraciones metafísicas acerca de la legitimidad de la explicación, ya que, como hemos visto, su propuesta pretende dejar de lado la metafísica.

Todo esto hace dudar seriamente de que el criterio propuesto por Baker y Burge de partir de un rango de explicaciones que resulten dignas de aceptación pueda resultar más que una expresión de deseos y no una estrategia, siquiera programática, para enfrentar con probabilidades ciertas de éxito el problema de la causación mental.

Por supuesto, Baker y Burge podrían alegar que la opción que favorecen es preferible aun cuando los problemas señalados sean reales, ya que al menos son tratables, mientras que intentar resolver el problema de la causación mental dentro del marco metafísico es imposible. Sin embargo, esto último no ha sido probado; no se ha presentado ningún argumento general que demuestre que esto no puede hacerse. Es innegable que el problema es de una complejidad extraordinaria y, quizás, insoluble. No obstante, dado el conjunto de deficiencias que hemos señalado sobre el intento de descartar la metafísica, no parece que tengamos a nuestra disposición una estrategia mejor para abordar el problema.

En el siguiente capítulo examinaremos otra propuesta cuyo objetivo es resolver el problema por medio de una estrategia basada en la escisión del *explanandum*: esta es la denominada 'estrategia del *explanandum* dual'.

CAPÍTULO VI: ESTRATEGIAS DEL *EXPLANANDUM* DUAL

1. *La estrategia del* explanandum *dual: algunos antecedentes*

Una de las estrategias utilizadas para lidiar con el problema de la exclusión causal/explicativa ha sido la denominada 'estrategia del *explanandum* dual' (o 'de los dos *explananda*'). En líneas muy generales, este enfoque 'intenta resolver la rivalidad explicativa sosteniendo que las dos explicaciones no comparten en realidad los mismos *explananda*' (Kim, 1991, p. 293).[112] Esta alternativa tiene antecedentes relativamente lejanos: en Von Wright (1971) se encuentra una exploración de la posibilidad de que las explicaciones 'teleológicas' y las explicaciones 'causales' no compartan los mismos *explananda*, en cuyo caso no se plantearía un problema de competencia o incompatibilidad entre ambas. Posteriormente, la estrategia ha sido empleada por Dretske,[113] en el marco del problema de cómo las razones, en tanto cierto tipo de sucesos mentales, contribuyen a explicar los efectos consistentes en outputs motores. El objetivo de Dretske fue encontrar una descripción de por qué las razones, en virtud de ser razones, explican causalmente las conductas que, por medio de su contenido, ayudan a explicar racionalmente. El problema planteado por Dretske puede ser caracterizado entonces como el problema de la *causación intencional* o *racional*, parte integrante del problema mayor de la causación mental (esto es, cómo los estados mentales pueden entrar en cualquier clase de relación causal, tanto como causa y como efecto). Más recientemente aún, Ausonio Marras (1998, 2000) ha revivido esta estrategia para hacer frente al problema general de la exclusión causal explicativa de las propiedades mentales.

El objetivo del presente capítulo es describir una de las versiones más recientes y elaboradas de esta estrategia, esto es, la presentada por Marras, la cual se inscribe en el marco de una consistente defensa del materialismo no reduccionista, y señalar lo que, a nuestro entender, constituyen limitaciones que arrojan serias dudas

[112] Esta caracterización de Kim es más general que alternativas más específicas y, quizás, menos preferibles. Sabatés (2001), como hemos visto en el capítulo anterior, sugiere que la posibilidad de salvar la causación de sucesos mentales por parte de otros sucesos mentales permitiría desarrollar una estrategia promisoria del *explanandum* dual. Vicente (2002) sigue una línea similar en la caracterización de esta estrategia.

[113] Su propuesta se encuentra en forma resumida en su (1990).

acerca de su posibilidad de conservar de manera plena la causación mental y la explicación psicológica en ese marco. Por último, analizaremos otra posición reciente que apela a la posibilidad de que existan explicaciones alternativas para los sucesos mentales, basadas en propiedades físicas y en propiedades mentales, pero limitando los poderes causales de estas últimas.

2. La crítica de Kim al enfoque de Dretske

Marras considera que su propio enfoque de los dos *explananda* evita los problemas que la versión de Dretske ocasiona; por esta razón, será conveniente comenzar describiendo brevemente la crítica de Kim a la estrategia en la forma que le da este último autor.

En su (1990), Kim expone las razones por las cuales la estrategia de los dos *explananda* desarrollada por Dretske no consigue solucionar el problema de la exclusión.[114] En líneas generales, el enfoque de Dretske puede ser descripto de la siguiente manera. En primer lugar, es necesario remarcar el carácter particular de los *explananda* de las explicaciones racionalizadoras en la concepción de este autor. Lo que se explica cuando se ofrece una 'razón', es una conducta o una acción, algo que el agente 'hace'. Dretske sostiene que un agente S hace A sólo si algún estado interno de S, S_i, *causa* un cierto *resultado* R (típicamente, un output motor) asociado con A. Hay una diferencia crucial, en el enfoque de Dretske, entre hacer A (la causación de R por parte de S_i) y R. La explicación de R, o la determinación de la causa de R, es algo muy diferente a la explicación de por qué S_i causó R. Si bien la neurobiología puede proporcionar una descripción exhaustiva de la etiología causal de R, incluyendo una descripción de *cómo* S_i causó R, no puede proporcionar una explicación de *por qué* S_i causó R. El trabajo de explicar esta relación causal es tarea indispensable y distintiva de la psicología; explicar por qué la estructura causal relacional S_i causó R es explicar por qué o cómo S_i ha llegado a unirse con R en el agente S, y el contenido de S_i explica, y es la causa de cómo este nexo causal ha llegado a ser como es.

[114] Si bien en su (1991) Kim desarrolla un análisis crítico mucho más minucioso de la propuesta de Dretske, en el cual examina con más detalle las mismas insuficiencias, para nuestros propósitos basta con el examen general aquí expuesto.

En opinión de Kim, la solución de Dretske al problema de la exclusión es un caso de lo que puede ser llamado estrategia 'de los dos *explananda*': las explicaciones biológicas y racionalizadoras no comparten los mismos *explananda*, por lo cual no existe competencia entre ellas ni exclusión mutua. Dado que las razones hacen racionales a las acciones en un sentido intencional, las explicaciones fisiológicas sólo pueden explicar movimientos corporales descriptos en términos puramente físicos; ya que las dos explicaciones parecen tener diferentes *explananda*, el conflicto potencial se desvanece.

Esta estrategia, advierte Kim, es atractiva: soluciona el problema por medio de la división del *explanandum*: la psicología se ocuparía de las acciones, estructuras causal-relacionales que adoptan la forma S_i-causa-R, y la neurobiología y las ciencias físicas estarían a cargo de los sucesos *simpliciter*, tales como R. Si bien Dretske no menciona en ningún momento el problema de la exclusión, señala Kim, ésta sería al parecer su respuesta si se le planteara explícitamente el problema. Sin embargo, advierte, es preciso señalar una característica extremadamente importante de esta estrategia: aun cuando funcione, es fundamentalmente dualista. Sostener que una acción, o fragmento de conducta, tiene dos o más aspectos distinguibles que requieren una explicación es una cosa; sostener, prosigue, que uno de los dos aspectos distinguidos puede ser explicado *solamente* recurriendo a una teoría intencional, o que sólo puede tener propiedades de contenido como causas, es algo muy distinto. Formulado de manera más explícita, el problema para esta estrategia se plantea de la siguiente manera: ¿son las causas que Dretske invoca, los supuestos *explananda* de las explicaciones psicológicas, entidades físicas? Si no lo son, entonces se ha adoptado un dualismo abierto. Pero si lo son, y son susceptibles de explicaciones causales físicas, entonces no podrán ser de utilidad para separar la psicología de la teoría física, y el problema de la exclusión surgirá nuevamente. Si no son susceptibles de explicaciones causales físicas, se está de nuevo ante una forma de dualismo psicofísico: existen entidades en el mundo físico para las cuales en principio no hay explicaciones causales físicas, sino sólo explicaciones intencionales.

Sin embargo, advierte Kim, Dretske es un fisicalista convencido, que no comulga con ninguna forma de dualismo. Lo que Dretske tiene en mente al sostener el carácter especial de las explicaciones psicológicas es, probablemente, no que tales

explicaciones son no físicas, sino más bien que tales explicaciones invocan propiedades *relacionales* de los estados internos de los agentes, mientras que las explicaciones neurofisiológicas de los 'resultados' de las causas refieren solamente a sus propiedades físicas *intrínsecas*. Pero esta distinción, si bien puede resolver un conflicto explicativo potencial entre psicología y biología, seguramente no resuelve el conflicto entre psicología y teoría física. La estrategia de los dos *explananda* por sí misma, concluye Kim, no solucionará el problema de la exclusión.[115]

3. Marras y una estrategia alternativa de los dos explananda

3.1. Realismo explicativo y exclusión explicativa

El principio de exclusión explicativa es uno de los principales aportes de Kim a la discusión sobre la causación mental. Suele advertirse que, si bien inaceptable en la manera en que es formulado por Kim, el principio de exclusión explicativa encierra una intuición plausible, por lo cual se hace necesaria una reformulación que evite los problemas de aquella y que a la vez preserve la intuición considerada válida. Sin embargo, el rechazo del principio y su reemplazo por una formulación no equivalente pero que conserve su espíritu no ha resultado una cuestión sencilla.[116]

En un artículo reciente Ausonio Marras (1998) desarrolla una propuesta en la línea de la última estrategia expuesta, y presenta una crítica tanto al principio de exclusión explicativa cuanto a dos de sus supuestos, esto es, el criterio de individualización de los sucesos y las explicaciones y el principio del realismo explicativo. En opinión de este autor, el principio de exclusión explicativa enunciado por Kim carece de apoyo apropiado y conduce a consecuencias inaceptables; sin embargo, considera que puede ser defendido un principio estrechamente

[115] Dretske (1995) admite que la estrategia de los dos explananda, tal como él la ha formulado, no consigue solucionar el problema: 'adoptar una estrategia del *explanandum* dual resuelve algunos problemas, no otros. En particular, no resuelve el problema de la exclusión explicativa' (p. 142). Sin embargo, advierte que es suficiente para resolver otros problemas, como el de la relevancia explicativa del contenido en tanto propiedad extrínseca de un suceso mental. Las insuficiencias de la variante de la estrategia propuesta por Dretske, por otra parte, no tienen por qué extenderse a todas las posibles versiones de ésta.

[116] El rechazo de los argumentos tendientes a probar la validez de la exclusión causal no es infrecuente entre los filósofos que defienden versiones del materialismo no reduccionista en relación con el problema mente-cuerpo. Entre otros, esta es la posición de Horgan (1997), quien ha afirmado que el razonamiento que conduce a la exclusión causal-explicativa, a pesar de la plausibilidad superficial que parezca tener, es profundamente erróneo, y que los materialistas no reduccionistas están ciertamente comprometidos con un robusto compatibilismo causal.

relacionado pero significativamente diferente. El rechazo y reformulación consecuente del conjunto de principios propuestos por Kim le permite desarrollar una defensa de la relevancia causal/explicativa de las propiedades mentales, en una variante de la estrategia de los dos *explananda*. Esta defensa de la relevancia causal de las propiedades mentales se inscribe dentro de una posición de defensa de los postulados del materialismo no reduccionista.

En los apartados que siguen describiremos con detalle la propuesta de Marras; nuestra reconstrucción de sus argumentos tenderá a mostrar cómo, en última instancia, su versión de la estrategia de los dos *explananda* requerirá de una respuesta sustantiva al problema de la relación entre propiedades mentales y propiedades físicas, y que esta respuesta no garantiza el rol requerido para las primeras dentro del materialismo no reduccionista.

Marras (1998) comienza su crítica analizando el principio del realismo explicativo propuesto por Kim. Este principio sostiene lo siguiente:

> RE: *C* es un *explanans* para *E en virtud* del hecho de que *c* mantiene con *e* alguna relación objetiva determinada *R* (Kim, 1988, p. 226).

La relación objetiva *R* es una relación que vincula a los sucesos en el mundo, y 'fundamenta' la relación explicativa entre las proposiciones que la componen. Cuando la explicación es una explicación causal (Marras advierte que ésta es la única clase de relación que le interesará en este trabajo) la relación objetiva *R* es la relación causal misma. Una explicación causal será verdadera o correcta en tanto *c* tenga lugar como efecto de una causa objetiva *e*.

La estrategia que Kim utiliza para mostrar como el realismo explicativo conduce a la exclusión explicativa es plantear un caso en el cual se ofrecen dos explicaciones de un suceso *e*, una de ellas en términos del suceso c_1 y la otra en términos del suceso c_2, y mostrar luego como, analizando las posibles interpretaciones de esta situación, se concluye que c_1 y c_2 deben estar relacionados y, dado RE, las explicaciones que aluden a esos sucesos no pueden ser distintas, o completas, o independientes. Dejando a un lado los casos de sobredeterminación, por otra parte excepcionales, ningún otro caso posible (casos de causas parciales, de

eslabones de una misma cadena causal, casos en los cuales una de las causas sobreviene o es reducible a la otra, etcétera) pueden considerarse como violaciones al principio de exclusión explicativa (PEE).

Marras está en completo desacuerdo con la afirmación de Kim acerca de que PEE constituye una restricción general plausible para cualquier modelo realista de la explicación. Por el contrario, sostiene, existe una robusta concepción realista de la explicación, diferente de la apoyada por RE, en la cual el principio de exclusión explicativa no tiene vigencia. En especial, considera que los argumentos de Kim en favor del PEE dependen de un criterio de individualización de las explicaciones que no sólo es implausible en sí mismo, sino que es también inconsistente con ciertos rasgos epistemológicos de la explicación que Kim mismo admite.

Bajo el principio del realismo explicativo propuesto por Kim, observa Marras, las explicaciones son individualizadas bajo un criterio estrictamente *extensional*: no importa cuáles sean las descripciones que correspondan a 'c_1' y 'c_2' en la afirmación de identidad '$c_1 = c_2$'; en tanto la afirmación de identidad sea verdadera, las dos afirmaciones, si bien lógicamente inequivalentes, proporcionan la misma explicación. Esta consecuencia es, para Marras, muy contraintuitiva. El contraejemplo que proporciona es el siguiente: puede estarse de acuerdo en que el terremoto causó el colapso del edificio si y sólo si el suceso que se informa en la página 5 de la edición de hoy del *Globe and Mail* lo hizo, dado que el último suceso *fue* el terremoto en cuestión. Pero, señala, si las explicaciones han de proveer comprensión, es al menos discutible que esas dos afirmaciones causales singulares provean la misma explicación de por qué el edificio colapsó. Dos afirmaciones causales singulares bien pueden ser extensionalmente equivalentes y carecer del mismo contenido explicativo. De hecho, observa, una afirmación causal puede ser verdadera y no ser en absoluto explicativa. Estas observaciones no constituyen una petición de principio, señala, ya que el propio Kim acepta el carácter *epistémico* de la noción de explicación: tener una explicación implica estar en posesión de una cierta clase de conocimiento; además rechaza la sugerencia de que la relación objetiva agota el contenido de una explicación. Debido a esto, prosigue, RE individualiza las explicaciones demasiado toscamente. Todo lo que requiere una forma razonable de realismo explicativo es la siguiente restricción:

> RE*: *C* es un *explanans* para *E sólo si* el suceso *c* realmente causó el suceso *e* (Marras, 1998, p. 443. Cursivas del autor).

Este principio es más débil que el planteado por Kim, observa Marras, ya que sostiene sólo que la relación causal entre los sucesos mencionados en las proposiciones *explanans* y *explanandum* es una condición meramente necesaria, no necesaria y suficiente, para la corrección de la explicación. La causación provee el 'fundamento objetivo' para la explicación causal en el sentido que una explicación de *e* en términos de *c* no será correcta a menos que *c* haya causado efectivamente *e*. Pero se requieren otras condiciones, epistémicas, para la corrección de la explicación, esto es, que la causa y el efecto sean descriptos en ciertos modos particulares y no en otros.

Marras acuerda, por otra parte, con la verdad de otros principios metafísicos admitidos por Kim:

> P1: Si *c* es [causalmente] suficiente para un suceso posterior *e*, entonces ningún suceso ocurrido en el mismo momento que *c* y completamente distinto de él es necesario para *e*.
> P2: Debe suponerse que todo suceso físico tiene una causa física (en la medida en que tenga una causa) y, en principio, una explicación física.

Y ambos principios implican conjuntamente:

> PCX: No puede haber más que una única *causa* completa e independiente de cualquier suceso (Marras, 1998, p. 444. Cursivas del autor).

Este principio es un principio de exclusión causal y, exceptuando la sobredeterminación, puede bien ser aceptado mientras se rechaza PEE. Y, a menos que se asuma la forma de realismo explicativo propuesta por Kim, la cual hace a la mismidad de la causa necesaria y suficiente para la mismidad de la explicación, no puede inferirse PEE a partir de PCX. La forma más débil de realismo explicativo (RE*), conjuntamente con P1 y P2, sólo apoyan la afirmación de que todo suceso que tiene una causa debe tener exactamente una causa completa e independiente.

3.2. La individualización de los explananda

Sugerir que la explicación no es extensional, sostiene Marras, es llamar la atención sobre el hecho de que la relación explicativa se produce entre sucesos *como un tipo*, o en la medida que ejemplifiquen una u otra propiedad. La forma canónica de una afirmación explicativa singular no es '*c* explica *e*', sino 'el que *c* sea *F* (o *qua F*) explica el hecho de que *e* sea *G*', y en donde los tipos *F* y *G* son identificados como la causa y el efecto respectivamente. Los debates recientes sobre la causación mental, prosigue, han enfatizado el hecho de que mientras la relación causal puede ser extensional, relacionando sucesos sin importar la manera en que son descriptos, una afirmación singular causal no llega a ser explicativa a menos que identifique la causa en términos de sus propiedades 'causalmente relevantes'. Marras considera que lo que hace a una propiedad causalmente relevante no es importante en esta discusión, sino que lo que importa es que una afirmación causal singular que pretende ser explicativa tendrá que individualizar la causa en términos de aquellas (entre sus indefinidamente muchas) propiedades que fueron relevantes para el efecto.

Este análisis se vincula con el problema de la exclusión de la manera que se expone a continuación. Supóngase dos explicaciones –una de ellas una explicación racionalizadora, y la otra una explicación neurológica- de un suceso singular B, por ejemplo, el que Jorge se levante del sillón:

E1: el que *c* sea una clase de suceso *R* causó *B*
E2: el que *c** sea una clase de suceso *N* causó *B*

c y *c** son sucesos que ocurren simultáneamente en Jorge. ¿Pueden E1 y E2 ser aceptados simultáneamente como explicaciones completas e independientes de *B*? La respuesta de Kim, basándose en el principio de exclusión, es negativa. No hay necesidad de analizar, sostiene Marras, las posibles maneras en que *c* y *c** están relacionados. Para refutar la afirmación de Kim es suficiente con identificar al menos una posible interpretación de E1 y E2 bajo la cual ambas serían aceptables como explicaciones completas e independientes de *B*, sin violar la clausura del mundo físico.

Supóngase entonces, sostiene Marras, de acuerdo con los ampliamente difundidos enfoques no reduccionistas de la relación mente-cuerpo, que $c = c^*$, pero que *R*, al ser múltiplemente realizable, es distinta de *N*. *B* debe ser considerado un suceso físico; sin embargo, debe recordarse que *B* tiene la estructura de un suceso de una clase determinada. Entonces, las dos explicaciones E1 y E2 deberían ser reformuladas de manera acorde, para reflejar el hecho de que *B* no es un suceso individual *simpliciter* sino un cierto *hecho* individual –el hecho de que *e* es un caso de un cierto tipo de sucesos-. Tan pronto como se conjuga esta consideración con la observación previa de que la relación 'causal' de la cual se habla en E1 y E2 es realmente la relación 'quausal', resulta claro que la forma canónica de las explicaciones del tipo ejemplificado por E1 y E2 es realmente:

c causó *e*, y esto fue en virtud del hecho de que *c* sea *F* y de que *e* sea *G*

Esta es la forma en la cual la expresión *c qua F* causó *e qua G* es entendida de manera standard en la literatura sobre la relevancia causal.[117]

En el caso objeto de discusión, es claro que no hay un solo tipo de suceso que el que 'Jorge se levanta del sillón' deba ejemplificar. Dada la naturaleza del ejemplo que se está examinando (una instancia de conducta intencional), podemos suponer razonablemente que el suceso en cuestión ejemplifica tanto las propiedades de ser una cierta clase de movimiento corporal cuanto la propiedad *distinta* de ser una cierta clase de acción intencional.

Sea entonces *A* la propiedad de levantarse intencionalmente del sillón, y *B* la propiedad fisiológica correspondiente. Entonces cada una de las dos explicaciones E1 y E2 puede ser reformulada canónicamente en dos formas:

[117] 'Lo que es entonces la cuestión en el problema de la causación mental no es tanto la *eficacia* causal de las propiedades mentales, sino más bien la *relevancia* de las propiedades mentales en la formulación de explicaciones causales. Llamemos a este tipo de relevancia explicativa *relevancia causal*, para distinguirla de otros tipos de relevancia explicativa que las propiedades mentales puedan tener (*i. e.*, para la formulación de explicaciones 'racionalizadoras'), y no dejemos que la *relevancia* causal sea confundida con la *eficacia* causal, la cual pertenece propiamente a sucesos. La objeción epifenomenista puede ser reformulada como sigue: en el no reduccionismo, el fisicalismo de casos, las propiedades mentales son causalmente irrelevantes. Llamo a esto 'objeción *epifenomenista de propiedades*', para distinguirla de lo que podemos haber llamado 'objeción *epifenomenista de sucesos*', previamente considerada. (*Epifenomenismo de sucesos*, entonces, es la concepción según la cual los sucesos mentales

E1(a): *c* causó *e*, y esto fue en virtud del hecho de que *c* sea *R* y de que *e* sea *A*
E1(b): *c* causó *e*, y esto fue en virtud del hecho de que *c* sea *R* y de que *e* sea *B*
E2(a): *c* causó *e*, y esto fue en virtud del hecho de que *c* sea *N* y de que *e* sea *A*
E2(b): *c* causó *e*, y esto fue en virtud del hecho de que *c* sea *N* y de que *e* sea *B*

Cada una de estas explicaciones, a su modo, pretende explicar por qué *e* ha ocurrido. El hecho de que *c* lo causó, claramente, subdetermina cuál de esas explicaciones es correcta. Lo que necesitamos saber es: ¿en virtud de que *c* causó *e*? Y para responder a esta cuestión se necesita saber, en primer lugar, cómo *e*, el suceso *explanandum*, es identificado como ejemplificación de un tipo determinado. Dado que, por hipótesis, *e* tiene tanto la propiedad *A* como la propiedad *B*, como se identifiquen tipos a los fines explicativos dependerá de razones contextuales –por ejemplo, si estamos interesados en explicar por qué una acción de tipo *A* ocurrió en esa ocasión, o por qué tuvo lugar un movimiento corporal de tipo *B*-. Si bien el suceso *explanandum* es uno y el mismo, los hechos *explanandum* son seguramente diferentes.[118]

Supongamos entonces que identificamos a *e* como una instancia de *A* –una acción de un determinado tipo-. ¿Causó *c*, en esa oportunidad, una acción de este tipo en virtud de poseer la propiedad racionalizadora *R* o en virtud de poseer la

---

son causalmente ineficaces; *epifenomenismo de propiedades* es la concepción según la cual las propiedades mentales son causalmente irrelevantes)' (Marras, 1994, p. 473. Cursivas del autor).

[118] La idea de que no se proporcionan explicaciones de sucesos considerados como un todo, sino de determinados aspectos del suceso, está ya presente en Hempel (1965). Hempel parte de la caracterización de lo que denomina 'hechos oracionales', los cuales se caracterizan por poder ser descriptos completamente por medio de una oración empírica; ejemplos de esta clase de hechos son 'la longitud de la barra de cobre *b* aumentó entre las 9,00 y las 9,01 de la mañana' y 'la extracción *d* dio una bolilla blanca'. Pero otra forma de concebir a un suceso único o particular es por medio de una descripción definida o un nombre de individuo, y no por medio de una oración que lo describe; ejemplos de tales 'sucesos concretos' (tal la denominación que les da Hempel) son 'el primer eclipse solar del siglo XX' y 'la erupción del Vesubio del año 79'. Hempel advierte que no es posible proporcionar una explicación completa de un suceso concreto, entendiendo por tal una explicación que de cuenta de todos los aspectos del suceso. Esto se debe a que un suceso concreto tiene infinitos aspectos diferentes, por lo cual no se lo puede describir en su totalidad y menos aún explicar completamente. De esta forma, señala, sólo tiene sentido pedir una explicación concerniente a lo que ha denominado 'hechos oracionales', mientras que en lo que se refiere a los sucesos concretos tiene sentido pedir explicaciones de sus aspectos o características, que son describibles por medio de oraciones (por ejemplo, 'que la erupción del Vesubio del año 79 duró tantas horas'). Estas observaciones de Hempel tienen un innegable parentesco con la idea de que pueden explicarse diferentes efectos apelando a distintos aspectos de la causa. Sin embargo, es necesario advertir una importante diferencia: Hempel no se compromete con una ontología determinada, en especial con una concepción particular de los sucesos y las propiedades. Más importante aún, Hempel no pretende que las explicaciones de aspectos parciales del suceso concreto son explicaciones del suceso total 'bajo una descripción'.

propiedad neurobiológica *N*? Que la respuesta sea una u otra dependerá de si la posesión de la propiedad *R* por parte de *c* fue causalmente relevante para la posesión de *A* por parte de *e*, o si la posesión de *N* por parte de *c* lo fue.

Marras no ofrece aquí una concepción de la relevancia causal (si bien observa que posee una),[119] pero considera que todo lo que necesita en ese contexto es que se acepte la siguiente condición –a la que considera no controvertible- para la relevancia causal:

> CR: donde *c* causa *e*, y donde *c* es un *F* y *e* es *G*, que *c* sea *F* es causalmente relevante para que *e* sea *G sólo si* el contrafáctico '$\neg Fc \rightarrow \neg Ge$' es verdadero (1998, p. 448).

Sobre la base de CR puede considerarse, sostiene Marras, a E1(a)-E2(b) como verdaderos sólo si los contrafácticos que implican son verdaderos. Correspondiendo cada uno a E1(a)-E2(b) tenemos:

(1a) $\neg Rc \rightarrow \neg Ae$
(1b) $\neg Rc \rightarrow \neg Be$
(2a) $\neg Nc \rightarrow \neg Ae$
(2b) $\neg Nc \rightarrow \neg Be$

En el enfoque no reduccionista que se ha asumido, prosigue Marras, estos contrafácticos tienen interpretaciones naturales de acuerdo con las cuales (1a) y (2b) resultan ser verdaderos, mientras que (1b) y (2a) resultan ser falsos.

Tenemos, entonces, dos explicaciones distintas y posiblemente verdaderas de porqué el suceso *e* ocurrió: E1(a) y E2(b). Nuestra fe en los esquemas explicativos de la neurofisiología nos persuade de la verdad de E2(b); y, suponiendo que nuestra creencia en la relevancia causal de las propiedades mentales está bien fundada, tenemos todas las razones para creer en la verdad de E1(a). Nótese, sin embargo, que E1(a) y E2(b) individualizan el suceso *explanandum* en diferentes modos –el primero como una cierta clase de acción intencional, y el segundo como una cierta clase de movimiento corporal-. *Relativamente a cómo sea individualizado el explanandum*, E1(a) y E2(b) no competirán el uno con el otro: E1(a) pretende explicar por qué un

[119] Describiremos esta propuesta en el apartado 3.4.

cierto tipo de acción ocurrió en una ocasión determinada, y E2(b) pretende explicar por qué cierto tipo de *movimiento corporal* lo hizo –y estos son hechos *explananda*-. Esto nos autoriza a decir, sostiene Marras, que *hay* un principio de exclusión explicativa PEE*, distinto de PEE, el cual puede muy bien ser verdadero:

> PEE*: No puede haber más que una explicación completa e independiente de un *explanandum* dado (1998, p. 449)

donde '*explanandum*' no quiere decir un suceso *simpliciter*, sino *el que un suceso sea de cierta clase*. Si tomamos a nuestros *explananda* como sucesos bajo una individualización por tipo –'bajo una descripción', en términos de Davidson- entonces la relación explicativa es no extensional; y bajo este esquema de cosas PEE es falso mientras PEE* bien puede ser verdadero.

Esta forma de tratar con el problema de la exclusión, sostiene Marras, es un caso de la estrategia denominada por Kim 'de los dos *explananda*'. Y si bien considera que las críticas que Kim ha elevado contra esta estrategia parecen sólidas en lo que respecta a la versión de la estrategia propuesta por Dretske, Marras afirma que no son aplicables a la versión que está proponiendo. Bajo la concepción davidsoniana de los sucesos, sostiene, *uno y el mismo* suceso físico puede ser un caso tanto de un tipo de acción como de un movimiento corporal. Por el contrario, en la formulación de Dretske surge el problema de si el *explanandum* 'S_i causó R', como diferente de R, es un suceso físico (en cuyo caso debería admitirse una explicación biológica aun cuando se acepte una intencional, con lo que el problema de la exclusión resurge), o si no lo es (con lo que no surge el problema de la exclusión, pero al costo de adherir a algún tipo de dualismo cartesiano).

3.3. La dualidad de los explananda *y el argumento de la superveniencia*

La posición de Marras, como hemos visto, requiere de una serie de supuestos complejos relativos a la naturaleza de la explicación, de la individuación de los sucesos y de la relación causal. Muchos de estos supuestos pueden ser discutidos y constituirse en bases para la réplica a los argumentos de Marras. Sin embargo, no objetaremos alguno de ellos en particular, sino que nos limitaremos por el momento

a llamar la atención sobre un aspecto que sólo es mencionado por el análisis de este autor.

Al describir el suceso davidsoniano 'Jorge se levantó del sillón', que ejemplifica dos propiedades (una propiedad racionalizadora y una propiedad neurobiológica), Marras afirma que no es necesario examinar las posibles maneras en que tales propiedades están relacionadas. En su opinión, como hemos visto, es suficiente encontrar al menos una posible manera en que ambas explicaciones sean aceptables como completas e independientes son violar la clausura del mundo físico. Esta suposición parece sumamente cuestionable. Marras acepta que la propiedad racionalizadora es distinta de la propiedad neurofisiológica y múltiplemente realizable por propiedades de esta última clase; sin embargo, estos rasgos alcanzan para excluir algunas posiciones con respecto a la relación entre tales tipos de propiedades (por ejemplo, algunas clases de dualismo, como el interaccionismo o el paralelismo, o la teoría de la identidad), pero no son suficientes para descartar otras, cuyas consecuencias para el plano explicativo no son menores. Un emergentista, por ejemplo, podría aceptar que la relación que vincula a las propiedades mentales con las propiedades físicas es una relación de realización (incluyendo la realizabilidad múltiple); sin embargo, aceptaría también la causación descendente, de lo mental a lo físico, como uno de los rasgos centrales de su posición.[120] En este último caso, es evidente que la aceptación de esta doctrina condiciona fuertemente lo que seremos capaces de afirmar acerca del rol que juegan las propiedades mentales en las explicaciones de la conducta.

Esta observación conduce a plantear una cuestión más de fondo. El análisis de Marras expuesto hasta el momento combina consideraciones ontológicas, relativas a la naturaleza de la relación causal y a la caracterización de los sucesos, con consideraciones epistemológicas concernientes a la naturaleza de lo que debe considerarse una explicación satisfactoria. Su objetivo principal ha sido mostrar cómo pueden ofrecerse dos explicaciones de un mismo suceso (conductual) sin que se plantee un problema de exclusión o competencia explicativa en los términos en que Kim propone. Ahora bien, hemos visto que, según las últimas (y más sólidas) formulaciones, el problema de la exclusión explicativa en el caso de los sucesos o

propiedades mentales se plantea más bien como una consecuencia del problema ontológico.[121] El interrogante que parece surgir naturalmente es, entonces, si podemos intentar resolver el problema de la exclusión explicativa para el caso de las propiedades mentales sin abordar el problema central, el de su ineficacia causal. En su (1998) Marras responde afirmativamente a esta pregunta, aunque de manera implícita. Sin embargo, tal análisis es desarrollado en su (2000). En esta nota crítica de *Mind in a Physical World*, Marras aplica algunas de las conclusiones expuestas en su (1998) al argumento de la superveniencia desarrollado por Kim. Las describiremos a continuación.

Marras considera al argumento de la superveniencia propuesto por Kim 'elegante y poderoso'; sin embargo, considera que hay maneras de enfrentarlo. Kim, observa, niega que el 'fisicalismo de casos' sostenido por Davidson sea suficiente para responder a la pregunta que formula el problema de la exclusión causal/explicativa, esto es: (P) 'Dado que todo suceso físico que tiene una causa tiene una causa física, ¿cómo es posible también una causa mental?' Al negar esta posibilidad, Kim debe suponer que, o bien que la relación causal no es extensional, o que los términos de la relación causal no son sucesos en el sentido davidsoniano (particulares espacio-temporales concretos e irrepetibles). Si se adopta una ontología de sucesos davidsonianos, prosigue, algunos de los cuales son tanto mentales cuanto físicos, parece ser suficiente para responder P; la causa mental de un suceso es idéntica como caso [*token identical*] con su causa física, salvo que se suponga que la relación causal no es extensional. Bajo esta interpretación, la afirmación singular causal '*c* causó *e*' (en la cual *c* y *e* son sucesos davidsonianos) es interpretada como una formulación elíptica para la afirmación singular 'quausal' '*c* causó *e en virtud de que c es F y e es G*', y en la cual *c* y *e* están relacionados por la relación compleja 'causa-en-virtud-de-<*F*,*G*>', donde <*F*,*G*> es un par ordenado de propiedades causalmente prominentes de la causa y el efecto respectivamente. De esta manera, continúa Marras, (P) puede ser reformulada en los siguientes términos:

[120] *Supra*, p. 50.
[121] *Cfr.* al respecto Sabatés (2001).

> (P*) Dado que todo suceso físico que tiene una causa es causado por un suceso en virtud de las propiedades físicas de éste, ¿cómo puede ser causado por el mismo suceso en virtud de sus propiedades mentales?

Marras observa que Kim, al negar que el fisicalismo de casos de Davidson sea suficiente para responder (P), quiere decir en realidad que no es suficiente para responder (P*). Esta lectura del problema, considera, se ajusta bien a la afirmación de Kim de que el problema de la causación mental involucra, en última instancia, la cuestión de la eficacia causal de las propiedades mentales. Y, desde esta perspectiva, el problema de la causación mental se transforma en cómo evitar el epifenomenismo de tipos, ya que el enfoque de Davidson es suficiente para evitar el epifenomenismo de casos.

Como alternativa, Kim puede estar suponiendo que si bien la relación causal es extensional, los términos de la relación son sucesos en el sentido de 'tropos' o 'ejemplificaciones de propiedades', y no en el sentido que les asigna Davidson. Si bien en el argumento de la superveniencia, observa Marras, Kim hace referencia a ejemplificaciones de *M* y de *P*, él mismo señala en otro trabajo que las ejemplificaciones de propiedades pueden ser interpretadas como tropos (una ejemplificación de *F* es un suceso),o, de manera alternativa, como algo (un suceso u objeto) que ejemplifica o posee *F*. Ambas formulaciones son, según Marras, equivalentes, ya que puede suponerse que los tropos causan lo que causan en virtud de sus propiedades constitutivas. De plantearse el problema en términos de tropos y sus propiedades constitutivas, este planteo sería el siguiente: dado que los tropos son individualizados por sus propiedades constitutivas, si las propiedades mentales son distintas de las propiedades físicas, ¿cómo un tropo físico que tiene una causa puede ser causado tanto por otro tropo físico cuanto por un tropo mental?

Provistos con las conclusiones precedentes, continúa Marras, hay que retornar al argumento de la superveniencia y considerar las relaciones causales *M-a-M** y *M-a-P**, las cuales, de acuerdo con la conclusión del argumento, son sólo aparentes en la medida en que surgen de procesos causales genuinos que ocurren de *P-a-P**. Sea entonces *c* un suceso davidsoniano que ejemplifica tanto *M* como *P*, y sea *e* un suceso davidsoniano que ejemplifica tanto *M** como *P** (y en donde *M* y *M** sobrevienen a partir de *P* y *P** respectivamente); puede preguntarse entonces:

1. ¿*c* causó *e* en virtud de que *c* sea *M* y *e* sea *M**, o *c* causó *e* en virtud de que *c* sea *P* y *e* sea *M**?
2. ¿*c* causó *e* en virtud de que *c* sea *M* y *e* sea *P**, o *c* causó *e* en virtud de que *c* sea *P* y *e* sea *P**?

Marras considera que las respuestas plausibles a estas preguntas son:

1. En virtud de que *e* sea *M** (en tanto sea identificada como un suceso mental de tipo *M**), *c* causó *e* en virtud de que *c* sea *M*, y no en virtud de que *c* sea *P*. Aun si *c* no hubiera sido *P* (sino, por ejemplo, *P'*, otra propiedad física en la base de superveniencia de *M*), habría causado igualmente un ejemplo de *M** (por ejemplo, un suceso *e'* que ejemplificara una propiedad física diferente, *P**', en la base de superveniencia de *M**). Pero si *c* no hubiera sido *M*, tampoco habría sido *P* (ya que *M* sobreviene a partir de *P*); cualquier cosa que *c* pudiera haber causado, entonces, no habría sido un ejemplo de *P** en la base de superveniencia de *M**, y por lo tanto no hubiera causado ejemplo alguno de *M**.
2. En la medida en que *e* sea identificada como de tipo *P**, *c* causó *e* en virtud de que *c* sea *P* y no en virtud de que sea *M*. Si *c* hubiera sido *M* sin ser *P* (sino, por ejemplo, *P'*, una base distinta de superveniencia de *M*), *c* no habría causado un ejemplo de *P** (sino, por ejemplo, un ejemplo de *P**').

Marras observa que la causación en-virtud-de depende generalmente de cómo la causa y el efecto son individualizados. La 'quausación' es una relación explicativa no extensional, y cuales descripciones puedan ser utilizadas para identificar a los términos de la relación depende de manera crucial de los intereses explicativos. Si lo que se quiere explicar es un ejemplo de un movimiento corporal, una explicación neurofisiológica que apele a causas neuronales será apropiada, pero si lo que se desea es explicar el mismo suceso como un caso de una acción intencional, se pretenderá una explicación en términos de una causa mental.

3.4. Un esquema de la dependencia psicofísica

Puede parecer extraño, *prima facie*, que el mismo esquema que ha sido de utilidad para salvar las explicaciones de la conducta que apelan a propiedades

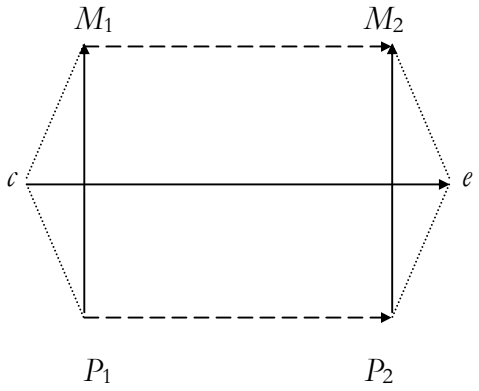
mentales en su presunta competencia con explicaciones que apelan a propiedades físicas sirva *también* para responder al argumento de la superveniencia. Esta objeción es anticipada por Marras, quien observa al respecto lo siguiente. Kim podría reclamar, señala, que este esquema confunde cuestiones metafísicas acerca de la causación con cuestiones epistemológicas acerca de la explicación causal; en otros términos, la pregunta acerca de qué propiedades de un suceso son *causalmente eficaces* con la pregunta acerca de que propiedades son *causalmente relevantes* (relevantes para explicaciones causales). Pero esto no es así, sostiene. En primer lugar, porque una propiedad no sería explicativamente relevante a menos que fuera causalmente eficaz. En segundo lugar, porque cuáles propiedades de un suceso son causalmente eficaces es una cuestión en sí misma relativa a qué *tipo* de efecto está en discusión. Propone que se considere una variante del ejemplo de Ernest Sosa: un sonoro disparo de arma de fuego bien dirigido causa simultáneamente una herida de bala y sordera en la víctima. En la concepción de Davidson de los sucesos que ha supuesto, hay una causa y un efecto, cada uno individualizado en términos de diferentes propiedades. Cuáles propiedades de la causa fueron causalmente eficaces para el efecto sólo puede ser determinado con referencia al *tipo* de efecto que se esté considerando: que la víctima se vuelva sorda, o que la víctima sufra una herida de bala. Marras considera que se debería sostener que si el efecto es un caso de lo último fue determinado porque la causa ha sido fue un disparo bien dirigido, mientras que si el efecto es un caso de lo primero fue determinado porque la causa haya sido un suceso sonoro.

Marras concluye el análisis de este problema sugiriendo que Kim está en lo correcto al afirmar que la eficacia causal de las propiedades mentales sobrevinientes no es básica, sino que 'surge', o 'depende de' la eficacia causal de las propiedades físicas subvenientes. Sin embargo, sostiene, dependiente o 'heredada', la eficacia causal es, con todo, eficacia causal; no habría razones para suponerla meramente 'aparente' o 'epifenoménica', ni tampoco para suponer que se está ante un caso de sobredeterminación, ya que una relación causal singular entre dos sucesos davidsonianos es descripta en diferentes niveles (los niveles sobreviniente y subveniente). Lo que la dependencia de los poderes causales de lo mental sobre los

poderes causales de lo físico implica no es que los primeros no sean reales, sino que no son primitivos y deben ser *explicados* en términos de los segundos.

Estas últimas afirmaciones requieren de una justificación, y esta justificación debe consistir en una respuesta definida acerca de la relación entre propiedades mentales y propiedades físicas. Esta tarea es afrontada por Marras en un artículo anterior (1997b).[122] Examinaremos ahora este aspecto de su propuesta.

En este artículo Marras analiza la necesidad de proveer una relación de dependencia psicofísica que rescate el status empírico y causal de las leyes psicológicas. Lo que se necesita en última instancia es, en su opinión, una explicación de la idea básica implícita en los puntos de vista fisicalistas de que todos los hechos y propiedades (incluyendo los hechos y propiedades mentales) dependen, o son determinados por, hechos y propiedades físicas. Su propuesta (esquemática, según su propia admisión) puede ser caracterizada a través del siguiente diagrama:



En la figura, c y e representan sucesos individuales (causa y efecto respectivamente), mientras que la flecha que los une representa la relación causal. M_1 y P_1 representan respectivamente la propiedad mental y la propiedad física ejemplificadas por c, y M_2 y P_2 representan la propiedad mental y la propiedad física (corporal) ejemplificadas por e. Las líneas oblicuas representan la relación de ejemplificación, las flechas verticales continuas representan la relación de dependencia/superveniencia, y las flechas segmentadas horizontales la relación legal (ya sea estricta o no estricta).

La idea de la dependencia metafísica que debe ser desarrollada puede ser esbozada a partir de las siguientes observaciones:

[122] Algunas de las ideas desarrolladas en este artículo se encuentran esbozadas en sus (1994 y 1997a).

1. Las propiedades mentales intencionales son construidas como propiedades funcionales de nivel superior, definidas en términos de su rol funcional característico en el sistema en el cual son ejemplificadas. Los estados que ejemplifican esas propiedades son típicamente construidos como disposiciones, donde la atribución de una disposición es la atribución de un estado causal, esto es, un estado que, bajo condiciones apropiadas, es causalmente responsable de determinados patrones de conducta.
2. Tal propiedad funcional *F*, en tanto es ejemplificable, es también realizable. Esto es, decir que *F* es realizable es decir que existe algún sistema de entidades de nivel inferior que es capaz de jugar el rol distintivo de *F*.
3. Debe suponerse también que las propiedades funcionales no son solamente realizables, sino que son *físicamente* realizables. Para cada propiedad funcional *F* hay un sistema de entidades físicas, y en última instancia, microfísicas, que son capaces de jugar el rol causal distintivo de *F*, como una consecuencia nomológica del tipo de entidades físicas que son y la manera en que se encuentran organizadas.
4. La estructura física subyacente, o sistema de entidades físicas que realizan la propiedad funcional *F* puede ser llamada *mecanismo físico* (uno entre muchos posibles). Este mecanismo implementa las generalizaciones implicadas por la atribución de la propiedad funcional y la disposición asociada con esa propiedad.

En opinión de Marras, si estas observaciones apuntan en la dirección correcta se trata del comienzo de una comprensión de cómo los poderes causales de las propiedades mentales sobrevinientes son determinados por, y dependientes de, los poderes causales de las propiedades físicas subvenientes. No obstante, esta estrategia debe afrontar, según Marras, algunas objeciones.

En primer lugar, podría objetarse que el esquema propuesto despoja a las propiedades causales de poderes causales propios, ya que éstos serían apropiados [*pre-empted*] por los poderes causales de las propiedades físicas que implementan los mecanismos de los cuales aquellas dependen. Marras encuentra a esta objeción algo bizarra; adscribir poderes causales a algo es típicamente postular la existencia de un mecanismo físico a través del cual pueda ejercer sus poderes causales. Por ejemplo,

cuando los científicos conjeturan que fumar causa cáncer, y buscan el mecanismo físico subyacente para sostener esta conjetura, resultaría bizarro sostener, una vez que se lo ha hallado, que la existencia de tal mecanismo muestra que realmente fumar no causa cáncer, sino que lo hacen las propiedades físicas del mecanismo subyacente. De manera similar, concluye, cuando las creencias y deseos causan la conducta lo hacen a través de las propiedades físicas subvenientes de los mecanismos que implementan las conexiones creencia-deseo-conducta.

Una segunda objeción podría ser la afirmación de que el status dependiente, de nivel superior, de las propiedades mentales, privaría a las leyes psicológicas de cualquier status causal genuino, ya que ellas serían verdaderas sólo debido a que ciertas leyes físicas que se aplican a los ejemplos relevantes de ejemplificaciones de propiedades mentales serían verdaderas. Marras repone que no hay razón para suponer que el status dependientes de las leyes psicológicas las hace menos causales; las leyes causales relacionan tipos de sucesos cuyas ejemplificaciones son pares de causa-efecto, y las leyes psicológicas no hacen esto menos que las leyes físicas, aun cuando lo hacen *porque* las propiedades mentales son ontológicamente dependientes de las propiedades físicas.

En tercer lugar, podría objetarse que la existencia de leyes físicas que se aplican a los mecanismos físicos subyacentes torna a las leyes y a las explicaciones psicológicas estrictamente superfluas. Cualquier interacción causal que involucrara a un suceso mental podría ser explicada directamente subsumiendo el suceso bajo leyes físicas, en virtud de las propiedades físicas del suceso. De esta manera, parecería redundante buscar una explicación del mismo suceso subsumiéndolo bajo leyes causales cuyo status causal es, en el mejor de los casos, derivado. Más aún, observa, la dualidad de explicaciones de un mismo suceso parece ser incompatible con el principio de exclusión explicativa.

Marras considera que hay varias respuestas a esta tercera objeción. En primer lugar, mientras la dependencia de las propiedades mentales sobre las propiedades físicas confiere status causal a las leyes psicológicas, no implica reducibilidad de lo mental a lo físico.[123] En su opinión, la relación de superveniencia preserva la

[123] En su (1993) Marras desarrolla detallados argumentos en contra de la tesis de que la superveniencia de las propiedades mentales a partir de las propiedades físicas pueda implicar la reducibilidad de las

distinción entre niveles de descripción y apoya la realizabilidad múltiple de las propiedades mentales. En consecuencia, las leyes psicológicas y físicas diferirán en alcance y generalidad; las leyes físicas y las leyes psicológicas, en una oportunidad dada, subsumirán clases no equivalentes de transacciones causales.

Por otra parte, las explicaciones son en general reconocidas como no extensionales y dependientes del contexto. La relación explicativa tiene lugar entre hechos, y estos hechos implican propiedades o tipos de sucesos, por lo que explicar por qué ha tenido lugar un determinado suceso en el momento *t* es explicar por qué un cierto *tipo* de suceso ha ocurrido en el momento *t*. Y, al solicitar una explicación de un suceso, debe preguntarse por qué *un tipo de suceso dado* ha ocurrido en esta oportunidad, y esta pregunta dependerá del aspecto del suceso que produzca perplejidad o interés, esto es, una cuestión dependiente del contexto.

Esta propuesta, observa Marras, es capaz de explicar cómo las explicaciones racionalizadoras pueden ser explicaciones causales. Difiere radicalmente del enfoque de Davidson en cuanto a que permite mantener la psicología dentro del dominio de las ciencias naturales, mientras que preserva su integridad como una ciencia especial. Permite reivindicar, en síntesis, las principales motivaciones que inspiran a las formas de materialismo no reduccionista: la creencia en la dependencia metafísica de lo mental sobre lo físico, junto con la creencia en la autonomía metodológica de la ciencia de la mente.[124]

3.5. Los límites de la propuesta

Esta extensa descripción de la propuesta de Marras ha tenido, entre otros objetivos, mostrar que, en última instancia, la estrategia del *explanandum* dual desemboca en la necesidad de explicar la causación psicofísica.

La propuesta de Marras plantea dos cuestiones de interés para analizar: 1) si el esquema propuesto salva completamente los poderes causales de lo mental, o los preserva únicamente en su propio nivel, con la consecuencia de que la conducta no

---

primeras a las segundas. Cabe observar, no obstante, que el análisis de las posibilidades del reduccionismo se basa en el modelo clásico de reducción propuesto por Nagel (1961), basado en la existencia de leyes puente bicondicionales. Quedan por determinar, por lo tanto, las posibilidades de modelos alternativos de reducción.

estaría determinada por lo mental, sino sólo por sucesos físicos; y 2) si el problema de la exclusión no se plantea también con respecto a la causación de un suceso mental por otro suceso mental. Estas dos cuestiones son centrales para determinar si el intento de Marras puede ser considerado exitoso. Las examinaremos sucesivamente.

La primera cuestión que analizaremos consiste en la posibilidad de que las propiedades mentales realizadas por propiedades sólo posean eficacia causal en su propio nivel; en otros términos, que no sean capaces de afectar causalmente a otras propiedades que no sean mentales.

A diferencia de, por ejemplo, el modelo de Jackson y Pettit, es evidente que el esquema propuesto por Marras no puede ser calificado de epifenomenista (o cuasi-epifenomenista), por dos razones obvias: en primer lugar, porque lo mental no depende causalmente de lo físico, sino que sobreviene a partir de él; en segundo lugar, porque si bien lo mental no tiene efectos sobre el mundo físico, es capaz de causar la ejemplificación de otras propiedades mentales. La propuesta de Marras, si es considerada exitosa, permitiría salvar las explicaciones de la ejemplificación de propiedades mentales (el hecho *explanandum*) que serían causadas por la ejemplificación de otras propiedades mentales.

Ahora bien, hemos visto que, para Marras, el contrafáctico $\neg Rc \rightarrow \neg Be$ (donde *R* es una propiedad racionalizadora y *B* es un movimiento corporal) es falso, dado que sólo una propiedad neurofisiológica es capaz de causar *B*; las propiedades mentales no causan la ejemplificación de propiedades físicas. Asimismo es notorio, en el esquema de la dependencia psicofísica propuesto por Marras, que no hay ninguna otra conexión entre propiedades mentales y propiedades físicas que no sea la relación de dependencia-realización. Las propiedades mentales sólo mantienen relaciones nomológico-causales con otras propiedades mentales. No hay causación descendente; las propiedades mentales no afectan causalmente a las propiedades físicas del suceso que ejemplifica ambas clases de propiedades. Si esto fuese así, lo que deseamos y creemos no tendría influencia causal sobre el mundo físico, en tanto

[124] Marras parece adjudicar un cierto carácter programático al esquema propuesto. Hasta donde hemos podido ver, no hay elaboraciones posteriores de la propuesta, por lo que parece razonable juzgarla aún reconociendo tal carácter provisorio y mejorable.

este último está determinado (causado) únicamente por propiedades físicas. Las propiedades puramente físicas de una acción (esto es, los cambios fisiológicos y mecánicos acaecidos en el cuerpo) sólo pueden ser explicados con referencia a propiedades puramente físicas.

¿En qué afecta esto a la propuesta de Marras? En escasa medida o de manera decisiva, dependiendo de los propósitos de la propuesta. Si la estrategia del *explanandum* dual limitara su alcance a salvar el rol explicativo y el poder causal de las propiedades (sin fijar una preferencia acerca del nivel en el cual estos poderes causales son ejercidos), podríamos considerar su intento como exitoso. Sin embargo, el objetivo general de la propuesta parece ser más ambicioso. El intento de Marras parece ser claramente construir una forma de salvar la causación mental y las explicaciones psicológicas de una manera enteramente compatible con el materialismo no reduccionista. Transcribimos al respecto una observación ilustrativa:

> Mi objetivo aquí es mostrar que hay una versión de materialismo no reduccionista que es capaz de sobrevivir a un importante tipo de objeción de la cual se ha dicho que el monismo anómalo de Davidson es vulnerable: la objeción *epifenomenista*. Esta objeción consiste en que el monismo anómalo de Davidson es incapaz de proveer una solución satisfactoria al problema de la causación mental –el problema de explicar cómo los sucesos mentales entran en relaciones causales con otros sucesos- (1994, pp. 466-7. Cursivas del autor).

Estas afirmaciones esclarecen el sentido de la propuesta de Marras. No se trata simplemente de mitigar las consecuencias indeseables de los argumentos de la exclusión (como el irrealismo o la irrelevancia explicativa), ni de salvar sólo parcialmente la causación mental. Se trata de mostrar cómo los argumentos de la exclusión son erróneos, y cómo lo mental puede conservar los poderes causales que ordinariamente le adscribimos. Si este es el objetivo fundamental de la propuesta, no podemos menos que concluir que, en el mejor de los casos, sólo se ha logrado parcialmente.

Hemos visto en capítulos anteriores que hay autores que conciben a lo mental como causalmente aislado.[125] Esto es, los sucesos o propiedades mentales no participan en ninguna clase de interacción causal, ni con otros sucesos o propiedades mentales o con sucesos o propiedades físicos; dependen (no causalmente) de sucesos físicos, vía superveniencia mereológica. En este modelo obviamente no puede existir sobredeterminación para los sucesos o propiedades mentales, ya que éstos sólo están determinados por una relación de superveniencia, y no causalmente. Ahora bien, en el modelo de Marras las propiedades mentales están determinadas (causalmente) por otras propiedades mentales y por propiedades físicas (por superveniencia/realización). Esta situación conduce a pensar si se trata de un posible caso de sobredeterminación.

Sabatés (2001) observa que una situación problemática de posible sobredeterminación entre dos determinantes (uno causal y otro por superveniencia) sólo puede plantearse si se admite un principio de exclusión ontológica que sostenga que no pueden existir dos conjuntos de propiedades completos e independientes cada uno de los cuales sea necesario para la ejemplificación de una propiedad singular. Tal principio es inevitable, observa, si se desea defender un principio de exclusión para los casos de causación de una propiedad mental a otra propiedad mental. Si este principio es aceptado, lo mental pierde completamente sus poderes causales; la base de superveniencia física de la propiedad mental excluye a la propiedad mental que es su supuesta causa.

Sin embargo, señala Sabatés, este principio no puede aceptarse sin alguna cualificación. No es posible admitir que cualquier par de relaciones de necesidad [*necessitation*] compitan o se excluyan mutuamente. Por ejemplo, que la estrella de la tarde sea visible hoy necesita [*necessitates*] que Venus sea visible hoy, pero que la estrella de la tarde sea visible hoy no parece evitar que la tarde sea clara hoy en Kansas por el hecho de que Venus sea visible hoy. Por lo tanto, para que el principio de exclusión de determinantes tenga posibilidades de ser plausible se debe admitir que las relaciones que pueden excluirse mutuamente son sólo aquellas que son tanto metafísica cuanto conceptualmente contingentes. Si una de las relaciones

[125] *Cfr.* la propuesta de Sabatés en el capítulo IV.

que presuntamente entran en competencia no se encuentra vinculada de manera contingente al efecto, la exclusión no tiene lugar. En este caso, señala Sabatés, la carga de la prueba recae sobre el defensor de la exclusión: es él quien debe mostrar que la relación de superveniencia entre lo mental y lo físico es una relación metafísica/conceptualmente contingente. Si esto no puede lograrse, podría mantenerse la causación de lo mental a lo mental. En este caso, se salvaría la mitad de la causación mental, dando lugar a una interesante posibilidad para una estrategia del *explanadum* dual.

Ahora bien, el esquema de dependencia psicofísica propuesto por Marras no apela solamente a una relación de superveniencia indeterminada entre lo mental y lo físico; como hemos visto, apela a la relación de realización física entre ambas familias de propiedades.

¿Debe ser entendida la relación de realización física como una relación contingente? Una caracterización standard de esta noción es la siguiente: un suceso *e* que sea F realiza el que *e* sea G sólo si i) *e* es F, ii) *e* es G, iii) para todo *e* es físicamente necesario que si *e* es F entonces *e* es G, y iv) el que *e* sea F explica el que *e* sea G. Sin embargo, esta caracterización no capta el sentido que esta noción posee para algunos funcionalistas, que sostienen que un estado físico puede realizar un estado funcional de manera contingente. El que exista una situación de competencia, y posible sobredeterminación, entre ambos determinantes dependerá entonces de cómo se conciba a la relación de realización física. Podemos considerar entonces a esta cuestión como abierta; si bien no hay argumentos decisivos para mostrar que se produce de manera efectiva una competencia entre determinantes, no hay tampoco razones para pensar que el peligro de la ineficacia causal de lo mental ha desaparecido completamente en virtud del esquema propuesto por Marras.[126]

[126] Hemos visto en el capítulo I que Kim, en las versiones más recientes de sus argumentos en favor de la exclusión causal (en especial, en el argumento de la superveniencia) parece inclinarse no sólo por afirmar la ineficacia causal de los sucesos mentales en relación con los sucesos físicos, sino por sostener la ineficacia causal completa (esto es, los sucesos mentales serían incapaces de causar nada, en particular, otros sucesos mentales). Recordemos al respecto su afirmación sobre la causación de un suceso mental por otro: 'en el caso de la supuesta causación *M*-*M**, la situación es casi como una serie de sombras proyectada por un automóvil en movimiento: no hay conexión causal entre la sombra del auto en un instante y su sombra en un instante posterior, siendo cada una de ellas un efecto del movimiento del auto' (1998, p. 45). El peligro de sobredeterminación no causal desaparece bajo esta descripción, pero al costo de concebir lo mental como completamente inerte.

La plausibilidad de la estrategia del *explanandum* dual propuesta por Marras, en síntesis, descansa en la plausibilidad de sus puntos de vista referentes a las relaciones entre propiedades mentales y propiedades físicas. Hemos intentado mostrar que el esquema de la dependencia psicofísica propuesto por Marras presenta aspectos dudosos y cuestiones sin resolver; en particular, si nuestra interpretación es correcta, no consigue asegurar un rol para las propiedades mentales en la producción de propiedades físicas, por lo que lo mental no jugaría un rol activo en la producción de la conducta. Si esto fuese así, su versión de la estrategia de los dos *explananda* no podría ser considerada exitosa.

A continuación examinaremos una propuesta que reconoce explícitamente que lo mental es incapaz de causar nada en el mundo físico, pero que es capaz de poseer efectos dentro de su mismo nivel. Esta propuesta, como veremos, parece situarse a mitad de camino entre las estrategias del *explanandum* dual que intentan mantener una relevancia causal y explicativa plena para las propiedades mentales y las variantes epifenomenistas o aislacionistas que niegan poderes causales a lo mental.

4. La independencia de la causalidad y la superveniencia

Thomasson (1998) comienza señalando los atractivos que algunas de las versiones del fisicalismo no reduccionista presentan, a la vez que advierte sobre el dilema que todas deben enfrentar: o las propiedades mentales tienen poderes causales, o no los tienen. Ambas alternativas conducen a consecuencias indeseables, aun cuando se acepten opciones teóricas aparentemente plausibles como el funcionalismo. Su propósito explícito es encontrar una solución al problema de la causación mental dentro del marco del materialismo no reduccionista, sin renunciar a ninguno de sus principios básicos. La tesis principal es que las dificultades que presenta el fisicalismo no reduccionista con respecto a la eficacia causal de las propiedades mentales son problemas meramente aparentes, resultado de una confusión fundamental entre causación y explicación. Una consideración cuidadosa de estas cuestiones conduce a una concepción de la causación mental que provee un rol causal apropiado para las propiedades mentales, sin renunciar a ciertos principios

comunes acerca de la causalidad y a la idea de que el mundo físico es ontológica y causalmente primario.

4.1. Una presunta confusión entre causación y explicación

Thomasson considera en primer lugar la necesidad de distinguir de manera inequívoca los diferentes tipos de relaciones que pueden subyacer a una explicación. En principio, entonces, señala que podrá existir más de una explicación de un suceso sin que esas explicaciones entren en competencia, y podrá haber más de una relación que involucre a los sucesos en cuestión sin que el suceso a explicar esté sobredeterminado causalmente.[127]

Como un ejemplo de explicaciones que apelen a diferentes relaciones subyacentes, pero que no compiten entre sí, Thomasson ofrece el siguiente caso: esas galletas están duras. Explicación 1: han sido sobrecocinadas. Explicación 2: sus moléculas están muy estrechamente ligadas. Estas explicaciones están muy lejos de competir entre sí, señala Thomasson; ambas pueden ser verdaderas sin que exista tensión entre ellas, y sin que deba preguntarse cuál es la única explicación verdadera. Ambas proveen una clase diferente de explicación. Mientras que la primera apela a factores causales en el pasado que producen el estado de cosas presente, la segunda recurre a las bases físicas subyacentes al estado de cosas en este momento.

Es extremadamente importante, continúa Thomasson, distinguir entre los tipos de explicación señalados y las relaciones subyacentes. Mientras que la primera relación es una relación de causación (típicamente, una relación diacrónica entre un suceso anterior y un suceso posterior que es producido por el primero), la segunda es una relación de 'determinación' (término con el que hace referencia a una relación sincrónica entre entidades de nivel superior y entidades de nivel inferior que contribuyen a que tengan las características que tienen). Un estado de cosas particular puede entonces ser tanto causado de una cierta manera como determinado a ser de esa manera. Estas relaciones no se excluyen mutuamente, ni la presencia de ambas constituye un caso de sobredeterminación, ya que sólo una es causal.

[127] Esta necesidad, subraya, estaba presente en la exigencia aristotélica de que una explicación completa de un suceso debería apelar a la totalidad de los tipos de causa.

Varias consideraciones son pertinentes con respecto a esta posición. En primer lugar, el ejemplo trae aparejadas las dificultades que implica el valerse de casos no científicos y luego extrapolar los resultados de este análisis a otros contextos. En este caso, podemos perfectamente aceptar como una explicación suficiente para nuestras necesidades pragmáticas la explicación de que las galletas han sido sobrecocinadas (por ejemplo, puede bastarnos saber que si deseamos evitar ese resultado es necesario un menor tiempo de exposición al calor) y, si sabemos algo sobre la estructura de la materia, puede ser apropiada la explicación acerca de la ligazón molecular.

Sin embargo, lo que puede valer como explicación satisfactoria para contextos no científicos puede carecer de validez para otros contextos. Es perfectamente plausible pensar que, en otro contexto en el que nuestras necesidades de comprensión sean mayores, las explicaciones podrían no ser competitivas sino complementarias. En efecto, conviene recordar que dos explicaciones alternativas de un mismo suceso pueden no ser competitivas simplemente porque ambas son explicaciones parciales. Una explicación más completa del suceso en cuestión, entonces, podría apelar tanto a la causación cuanto a la relación micro-macro (determinación, en los términos de Thomasson). De esta forma, una explicación más completa haría referencia tanto a la naturaleza de la ligazón molecular como a la temperatura a la que la sustancia ha sido sometida.

No obstante, para evaluar correctamente la plausibilidad de esta última posibilidad se hace necesario plantear los nexos entre las relaciones objetivas que sirven de fundamento a la explicación. Una cuestión que parece crucial y que Thomasson no plantea es la siguiente: ¿cada una de estas relaciones es suficiente para la producción de la propiedad? Si lo son, no parece posible evitar la sobredeterminación; si no lo son, si son individualmente necesarias y conjuntamente suficientes, entonces lo que tenemos en caso de encontrarnos con dos explicaciones son dos explicaciones parciales. Veremos, al describir la posición de Thomasson con respecto a la causación mental, que esta cuestión no tiene una respuesta satisfactoria.

Provista de las distinciones precedentes, Thomasson examina el argumento de Kim en contra del materialismo no reduccionista expuesto en su (1993a).[128] Contrariamente a la conclusión de Kim, Thomasson no considera que la existencia de dos explicaciones del mismo suceso mental M^* creen una tensión epistémica. Siguiendo su esquema inicial, considera que la explicación 'M^* existe porque M la causó' expresa una relación causal, mientras que 'M^* existe porque es realizada físicamente por P^*' expresa una relación de constitución material y determinación entre M^* y P^*. Dado que estas explicaciones no compiten, no existe tensión generada por ellas. [129] La plausibilidad del argumento de Kim, observa Thomasson, deriva 'sólo de nuestra errónea tendencia a confundir causación y determinación, o de pensar en la determinación como una relación causal' (p. 186).

La solución que Thomasson propone está basada, señala, en el modelo estratificado del mundo. Un principio común a las concepciones estratificadas es que la causación ocurre sólo dentro de cada nivel; no hay causación ascendente o descendente. Los niveles, entonces, están conectados por relaciones de determinación, dependencia y constitución material. En tal visión estratificada, las propiedades mentales están ubicadas entre las propiedades de nivel superior, y dependen de ejemplos particulares de propiedades físicas que las realizan. Cada ejemplo de una propiedad mental es en última instancia exclusivamente dependiente y completamente determinada por propiedades físicas (lo que asegura la primacía de lo físico).[130] Sin embargo, las propiedades mentales, en virtud de sus rasgos semánticos, o cualitativos o de su dependencia del contexto social, no pueden ser identificadas ni como tipos ni como casos con las propiedades físicas de las cuales dependen.

Aun suponiendo, prosigue Thomasson, que las propiedades y relaciones causales más fundamentales se encuentren en el nivel microfísico, pueden existir

[128] El argumento de la superveniencia de Kim, que hemos expuesto en el capítulo I, puede considerarse equivalente en lo que respecta a la ineficacia causal de las propiedades mentales.
[129] Es en este sentido en que es posible interpretar la posición de Thomasson como una variante de la estrategia del *explanandum* dual: tendremos dos explicaciones no competitivas de la misma propiedad mental, una de ellas causal y la otra por determinación no causal. Los sucesos humanos, como veremos, admitirán asimismo dos niveles explicativos compatibles entre sí.
[130] Thomasson observa, sin embargo, que una propiedad mental quizás no dependa *solamente* de la ejemplificación de la propiedad física que es su base, ya que puede estar también parcialmente determinada por aspectos del contexto físico y aún social (contexto que, a su vez, será en última instancia completamente dependiente del contexto físico).

objetos y propiedades de nivel superior que dependan pero que no sean reductibles a los objetos y propiedades fundamentales. En tal caso, podrían existir relaciones causales en los niveles superiores que dependan de las relaciones causales de los niveles inferiores sin ser reductibles a ellas. Como ejemplo de esta clase de relaciones, Thomasson propone el siguiente: la especulación de los inversores produce el colapso de los mercados. No es probable que creamos en esta historia causal a menos que existan también relaciones basadas en interacciones físicas fundamentales entre los especuladores y los mercados (interacciones tales como llamadas telefónicas o envíos de faxes). Las relaciones causales de niveles superiores son en sí mismas *dependientes* de relaciones causales fundamentales, lo cual disminuye la impresión de que poderes causales nuevos y misteriosos tienen lugar en tales niveles. Las relaciones causales nuevas, así como los sucesos invocados en ellas, son, en última instancia, dependientes de entidades en el nivel físico básico.[131]

La pregunta '¿tiene lo mental poderes causales?' puede entonces ser respondida afirmativamente por el materialista no reduccionista. Lo mental tiene poderes causales distintos, dependientes pero no reductibles. Esta solución, en opinión de Thomasson, evita el peligro del epifenomenismo, y también satisface el *dictum* de Alexander, ya que concede realidad a lo mental a través de la asignación de poderes causales.

El hecho de que las propiedades mentales sean determinadas por propiedades físicas y causadas por otras propiedades mentales, en opinión de Thomasson, no implica que se esté ante un caso de sobredeterminación global. Habiendo descartado que se trate de un caso de sobredeterminación causal, ya que sólo una de las relaciones es causal, debe desecharse igualmente la posibilidad de que se trate de una sobredeterminación no causal –M^* es causado por M y determinado por P^*-. Pero M^* no está más sobredeterminado, sostiene Thomasson, que la dureza de la galleta debida a la sobrecocción y a la estrecha ligazón de sus moléculas. No debería confundirse, observa Thomasson, suficiencia y causación: 'P^* y M son

[131] Thomasson no menciona el modelo de causación superveniente propuesto por Kim, que parecería ajustarse a la idea de dos explicaciones no competitivas del mismo suceso. Una razón posible para esto es que, como hemos visto al analizar la propuesta de Marras, este modelo no garantiza (en opinión del propio Kim) que los poderes causales de los sucesos sobrevinientes sean algo más que poderes meramente aparentes o epifenoménicos.

"suficientes para" *M** en diferentes maneras: uno es suficiente para causarlo, el otro es su base de suficiencia física' (p. 190).

Ahora bien, ¿qué significa esta última afirmación? Recordemos que la ejemplificación de la propiedad *M** está completamente determinada por la ejemplificación de la propiedad *P**. Se hace aquí pertinente el análisis que hemos realizado con respecto al ejemplo provisto por Thomasson. La relación entre *M**, por una parte, y *M* y *P**, por la otra, no puede tratarse de un caso de sobredeterminación debido a que, en este último tipo de casos, cada determinante es suficiente para la producción del suceso; en este caso, *M* podría ser suficiente para producir *M**, en ausencia de *P**, y esto violaría el principio de que cada propiedad mental está completamente determinada por una propiedad física (*P**). Tenemos entonces dos opciones: o la ejemplificación de la propiedad mental *M** está completamente determinada por las propiedades físicas subyacentes, o no lo está. En el primer caso, la propiedad física *P** subyacente sería *suficiente* para la ejemplificación de la propiedad mental, y no haría falta ninguna otra propiedad para su ocurrencia. Pero esta alternativa dejaría a la propiedad *M* causalmente ociosa.

La última alternativa que parece quedar disponible es que tanto *M* cuanto *P** sean necesarias para la producción de *M**. Si *P** es necesaria, pero no suficiente para la ejemplificación de *M**, quizás podría decirse que tanto *P** como *M* son necesarias y conjuntamente suficientes para la ejemplificación de *M**. En el último caso, parte de las características de la propiedad mental en cuestión serán fijadas (vía causación) por otra propiedad mental. Pero esto es contradictorio con la idea de que cada propiedad mental es exclusivamente dependiente y completamente determinada en última instancia por propiedades físicas. No se comprende, por lo tanto, por qué Thomasson insiste en que no existe ninguna clase de competencia, exclusión o complementariedad entre ambos determinantes.

Estas objeciones, como veremos, son menores en comparación con los límites que Thomasson reconoce a su propuesta.

4.2. Causación mental sin causación psicofísica

La solución propuesta, sostiene Thomasson, tiene muchas virtudes, ya que nos posibilita preservar las propiedades mentales en su naturaleza distintiva, les provee de un rol causal, evita los problemas de la sobredeterminación o la violación de la clausura causal de lo físico, y preserva la visión científica que concibe al mundo físico como fundamental.

Sin embargo, esta propuesta no carece de limitaciones, que Thomasson plantea con claridad. En principio, impide la aceptación de las afirmaciones comunes que implican la causación de lo mental a lo físico y de lo físico a lo mental. No sería literalmente verdadero, entonces, que la ingestión de una droga *causa* el alivio del dolor de cabeza. Lo que realmente tendría lugar sería una cadena causal en el nivel fisiológico (la ingestión de un medicamento actuaría causalmente sobre la zona afectada), y una relación de determinación entre el nivel fisiológico y el nivel psicológico (la modificación funcional en la zona afectada determinaría la desaparición del dolor).

La situación es diferente con los presuntos casos de causación de lo mental a lo físico. Existen varias razones, sostiene Thomasson, que hacen más fácil la aceptación de la inexistencia de tales casos. En primer lugar, si bien el modelo expuesto no deja espacio para la causación mental descendente de sucesos físicos, puede haber causación mental de sucesos de nivel superior como dimisiones, guerras o colapsos de mercados. Esos sucesos culturales significativos están basados en el mundo físico, pero no son sólo entidades físicas fundamentales.

En segundo lugar, Thomasson observa lo siguiente:

> Limitar los poderes causales de lo mental a suceso de nivel superior, descartando efectos al nivel físico básico, puede no parecer tan malo si recordamos las alternativas. La elección parece darse entre las siguientes concepciones no reduccionistas: 1) elegir el funcionalismo e identificar ejemplos de propiedades mentales con sus bases físicas, con el riesgo de perder los rasgos cualitativos, semánticos o de contenido, esenciales de lo mental, o 2) continuar siendo no reduccionista tanto en el nivel de tipos cuanto en el de casos y a) tratar lo mental como un epifenómeno (claramente una opción menos atractiva para aquellos que toman las intuiciones acerca de lo mental seriamente), o b) aceptar la causación mental-físico y sus cabales concomitantes de sobredeterminación y violación de la clausura causal de lo físico, o c) postular la causación mental, pero limitándola a los sucesos de nivel superior (p. 192).

El argumento de Kim, concluye Thomasson, no nos da razones para abandonar el fisicalismo no reduccionista. Si bien resta desarrollar una descripción de las relaciones causales de nivel superior y un análisis de las relaciones de determinación y causación, observa, el proyecto no reduccionista no está muerto. Renunciar a él debido al temor de que no exista una réplica consistente al dilema de la causación mental sería abandonar prematuramente una teoría prometedora e interesante.

Las alternativas planteadas por Thomasson no resultan tan poco atractivas cuando se advierten las implicaciones reales de su propuesta. La ausencia de causación de un estado mental por un estado físico no implica ningún problema, si recordamos que los estados mentales están determinados no causalmente por sucesos físicos. Ahora bien, si la causación de los estados físicos por estados mentales está vedada, ¿qué significa realmente afirmar que 'puede haber causación mental de sucesos de nivel superior tales como renuncias, guerras, colapsos de mercados de capitales, manifestaciones de amor y actos de vandalismo' (p. 191)? Lo que significa, al parecer, es que el mundo de los sucesos humanos se estructura en dos secuencias causales paralelas, estando cada suceso mental unido a un suceso físico por una relación de determinación, pero sin que exista ninguna forma en la cual la cadena causal de sucesos mentales influya sobre los sucesos físicos. Si esta interpretación es correcta, el mundo de lo humano resulta casi más extraño que si se negara a los sucesos mentales todo poder causal (como ocurre en el aislacionismo).[132] Observaríamos un mundo en el cual ocurren sucesos físicos sin la injerencia o intervención de sucesos mentales, los cuales son causalmente inertes en ese mundo. Viviríamos en un mundo al cual creemos (hasta cierto punto) controlar, pero que en realidad se mueve de manera autónoma, guiado, es de suponer, por leyes puramente físicas. Esta posibilidad, podemos concluir, difícilmente resulte aceptable para un materialista no reduccionista que, como afirma Thomasson, tome seriamente las intuiciones acerca de lo mental.

Hemos analizado en este capítulo algunas posiciones que recurren a la división del *explanandum* en busca de una solución al problema de la exclusión causal. Podemos concluir que esta estrategia enfrenta problemas de diversa clase, si bien no puede afirmarse que ninguna de sus variantes pueda constituir un progreso con respecto al problema. En el capítulo siguiente examinaremos un intento de sostener una forma de dualismo más fuerte que el simple dualismo de propiedades, basado parcialmente en la presunta ganancia explicativa que tal doctrina posibilitaría.

[132] *Supra*, p. 118.

CAPÍTULO VII. 'LA OSCURA CAVERNA DEL DUALISMO'[133]

1. El abandono del dualismo: ¿una situación irreversible?

Resulta frecuente hallar, entre los filósofos dedicados al problema mente-cuerpo, menciones a las graves dificultades que plagan a cualquier forma de dualismo que pretenda sostener algo más fuerte que la existencia de propiedades mentales dependientes de estados físicos, esto es, un 'modesto' dualismo de propiedades. Dicho en otros términos, a cualquier dualismo que defienda la existencia de 'entidades' mentales cualitativamente diferentes de la materia física y cuyo estudio se encuentre fuera de las posibilidades de la ciencia. No todas estas dificultades aparentan ser insalvables;[134] argumentos conocidos contra el dualismo, como su presunta 'extravagancia ontológica' (esto es, la complicación innecesaria de la ontología), no parecen ser concluyentes. Sin embargo, la dificultad de explicar la interacción causal entre sustancias o entidades de naturaleza cualitativamente diferente, problema que aquejó a Descartes y que fue señalado por sus propios contemporáneos, parece ser un problema irreductible.[135] Posteriormente, a la dificultad de explicar tal interacción se agregó la incompatibilidad con principios fundamentales de la física, como el principio de conservación de la energía. Los intentos relativamente recientes de defender una posición interaccionista, como el de Popper y Eccles (1977) no parecen haber mejorado mucho las perspectivas del dualismo en estos respectos. Una prueba de las dificultades con las cuales estos autores se enfrentaron al intentar sostener su interaccionismo se refleja en las tentativas de Popper, que aparecen en las discusiones finales, de proponer alguna hipótesis que permita evitar la violación de la primera ley de la termodinámica o principio de conservación de la energía.

Sin embargo, sería un error suponer que el dualismo ha sido completamente desterrado del campo de las teorías sobre el problema mente-cuerpo y, en particular,

---

[133] 'Para muchos de nosotros, el dualismo es un territorio desconocido, y tenemos poco conocimiento de que posibilidades y peligros acechan en esta oscura caverna' (Kim, 1998, p. 120).

[134] *Cfr.* Bechtel (1988), Bunge (1981), Churchland (1988), para una exposición general de los argumentos en favor y en contra del dualismo.

sobre la causación mental.[136] Pueden hallarse defensas de posiciones dualistas en compilaciones sobre temas de filosofía de la mente en las cuales predominan de manera abrumadora las perspectivas materialistas.[137] No es nuestro objetivo aquí analizar una teoría dualista general de la relación mente-cuerpo; no obstante, sí resulta de interés examinar un intento de articular una forma de dualismo para responder al problema de la aparente ineficacia causal de las propiedades mentales, tentativa que lleva a cabo E. J. Lowe en un artículo reciente (1999). Esta tentativa presenta el interés, para nuestros propósitos, de apelar a las presuntas ganancias explicativas que tal forma de dualismo supondría. En lo que sigue, expondremos las principales tesis del intento de solución de Lowe, para intentar mostrar luego que muy difícilmente puede ser considerado exitoso. En particular, la propuesta de Lowe parece tropezar con dificultades análogas a las enfrentadas por prominentes dualistas del pasado, y no parece constituir una mejora sustantiva con respecto a ellos.

2. El 'dualismo naturalista' de Lowe

Las afirmaciones que se transcriben a continuación constituyen el marco general del análisis de Lowe. En su opinión resultan de aceptación obligada, pero no parecen constituir un conjunto consistente:

(1) El sí mismo [*self*], si bien físicamente corporizado [*embodied*], no se identifica con ningún cuerpo físico ni con ninguna parte de tal cuerpo.
(2) El sí mismo es por su propia naturaleza un agente, algo que es naturalmente capaz de efectuar acciones intencionales, algunas de ellas con resultados físicos.
(3) Cada suceso físico tiene un conjunto de causas completamente físicas que son colectivamente causalmente suficientes para la ocurrencia de ese suceso (y raramente, si es que en algún caso, un suceso físico es sobredeterminado causalmente) (p. 225).

Contradiciendo la opinión de muchos filósofos, que al considerar inconsistente este conjunto de afirmaciones se ven en la necesidad de rechazar

[135] Por supuesto, esto vale sólo para aquellas formas de dualismo de sustancias que pretendan mantener la eficacia causal de lo mental (las variantes del interaccionismo serían el ejemplo privilegiado de esta posición); no se aplica a aquellas que no pretenden esto (el paralelismo o el epifenomenismo).
[136] Kim (1998), menciona como ejemplos recientes de filósofos que defienden al dualismo a R. Swinburne (*The Evolution of the Soul*, Oxford, Clarendon, 1986), John Foster (*The Inmaterial Self*, London, Routledge, 1991, y W. D. Hart (*The Engines of the Soul*, Cambridge, Cambridge University Press, 1988).

alguna de ellas, Lowe afirma que tal conjunto es perfectamente consistente. Dado que la tesis (3) goza de una aceptación muy amplia, Lowe argumenta exclusivamente en favor de las dos primeras, mucho más controvertidas.

Una persona o sí mismo, para Lowe, es un ser que puede tener pensamientos acerca de sí, pensamientos tales como 'siento calor' o 'mido un metro con ochenta centímetros'. Pero una persona o sí mismo, si bien es físicamente corporizado, nunca se identifica con su cuerpo físico ni con ninguna parte de él (tesis (1)). No estamos en absoluto inclinados, observa Lowe, a considerar que nuestros cuerpos o nuestros cerebros sean el sujeto de nuestros pensamientos en primera persona. Y dado que ni nuestro cuerpo ni ninguna parte de él es el sujeto de nuestro pensamiento, se sigue que el sí mismo no es idéntico al cuerpo ni a ninguna de sus partes. Sin embargo, prosigue, sería un error pensar que de esto se sigue necesariamente que el sí mismo debe ser identificado con algo no físico, tal como un espíritu o alma o 'ego cartesiano'. El sí mismo puede ser una cosa física sin ser idéntico al cuerpo o al cerebro.

Como una parte de esta primera tesis, Lowe afirma que los estados mentales *no son* estados físicos. Entre otras razones, Lowe sostiene que un estado físico, por su propia naturaleza, es un estado cuya posesión por parte de una entidad introduce alguna diferencia real en al menos parte del espacio que esa entidad ocupa. Por el contrario, un estado mental no tiene connotación espacial alguna. De hecho, prosigue, las condiciones de identidad de los estados mentales parecerían ser totalmente distintas de las de los estados físicos. Consecuentemente, concluye, la tesis de que los estados mentales 'son sólo' (o idénticos a) estados físicos es sencillamente ininteligible. Lamentablemente, agrega, una generación entera de filósofos ha considerado erróneamente a esta tesis ininteligible como una verdad profunda que sólo ahora nos ha revelado el avance de la ciencia.

Lowe no argumenta en detalle acerca de estas afirmaciones, sino que remite a sus (1989 y 1996)[138] para un análisis completo de estas cuestiones. No discutiremos aquí los dos aspectos de la tesis (1) expuestos por Lowe, aunque, cabe decir, ambos

[137] *Cfr.* Hart (1994).
[138] Lowe, E. J. (1989), *Kinds of being: A Study of Individuation, Identity and the Logic of Sortal Terms*, Oxford, Blackwell; (1996), *Subjects of Experience*, Cambridge, Cambridge University Press.

son fuertemente controvertidos, en particular el argumento en favor del segundo.[139] Los aceptaremos *for the sake of the argument*.

La tesis (2) –la idea de que el sí mismo es, por su propia naturaleza, un agente- también requiere de una defensa. La caracterización del sí mismo como algo que necesariamente es capaz de autorreferencia, observa Lowe, es también una caracterización de algo que es capaz de ser un agente, ya que la autorreferencia es una clase de acción intencional. Más aún, observa, el desarrollo de la autoconciencia está ligado con el desarrollo de la conciencia de otros, y ambos están necesariamente ligados con el desarrollo de capacidades comunicativas, lingüísticas o no. Si esto fuese así, sostiene, no podría existir un sí mismo que fuese constitucionalmente incapaz de comunicarse con otros sí mismos a lo largo de su vida. Por último, Lowe observa que otra razón para pensar que el sí mismo es capaz de acciones intencionales es que sólo un ser capaz de tales acciones podría desarrollar un concepto de causación, y que poseer tal concepto es una condición necesaria de la autorreferencia y de la mismidad.

Una vez presentadas las defensas de las tesis (1) y (2), Lowe examina la posibilidad de que, conjuntamente con la tesis (3), constituyan un conjunto de afirmaciones inconsistente. Parte de la tesis (1), como hemos visto, consiste en afirmar que los estados intencionales no son estados físicos, mientras que la tesis (2) implica que los estados intencionales pueden ser causas de sucesos físicos (dado que el concepto de acción intencional es causal). Si a esto agregamos la tesis (3), esto es, que todo suceso físico tiene un conjunto de causas completamente físicas que son colectivamente causalmente suficientes para su ocurrencia, parecería que el conjunto de afirmaciones es inconsistente, ya que los estados no físicos son y no son causas de sucesos físicos. Sin embargo, observa Lowe, esta conclusión es incorrecta, ya que ignora la transitividad de la causación. La cuestión clave a tener en cuenta, señala, es que si x es causalmente suficiente para y e y es causalmente suficiente para z, entonces, por transitividad, x es causalmente suficiente para z, aun cuando esto no implica que z es causalmente sobredeterminada por x.[140]

[139] Ya en los escritos de los primeros teóricos de la identidad (en particular en los de J. J. C. Smart) se encuentran respuestas a este tipo de objeción.

[140] Manuel Comesaña ha objetado que el punto de partida de la argumentación de Lowe se basa en una redefinición estipulativa subrepticia de 'causa'. Observa que, en el uso común, todo suceso tiene una

Es posible que la tesis (3) sea verdadera y que, no obstante, un suceso físico *P* tenga, entre sus causas, a un estado no físico *M*, sin que esto implique la sobredeterminación causal de *P*. Esto es así, observa Lowe, debido a que *M* puede tener un conjunto de causas completamente físicas que son conjuntamente suficientes para su ocurrencia. Si M es una causa de P, entonces todas las causas físicas de *M*, por la transitividad de la causación, son también causas físicas de *P*. Por lo tanto, concluye Lowe, (1) y (2) no son inconsistentes con (3), sino con una afirmación mucho más fuerte:

(4) No hay ningún suceso físico que tenga una causa no física (*ibid*., p. 230).

Ni la tesis (4) ni la tesis (3) (mucho más débil que la anterior) han sido fuertemente confirmadas por la evidencia empírica. La tesis (4), agrega, es más bien un 'artículo de fe' para algunos filósofos. Sin embargo, hay presunciones en favor de (3), ya que la ciencia moderna nos alienta a creer que el universo es un sistema causalmente cerrado cuyo origen fue completamente físico. Pese a esto, observa Lowe, esta presunción no obliga a desechar la posibilidad de que, en algún momento de la evolución del universo, puedan haber surgido sucesos o estados no físicos, junto con sujetos de esos sucesos o estados. Esta idea no involucra elementos de sobrenaturalismo, por lo que no hay razón para despreciarla como 'fantasmática' [*spooky*].

3. El principio de clausura causal del mundo físico

Es evidente que la defensa del principio modificado de clausura causal del mundo físico es crucial en el argumento de Lowe, ya que de él depende que el conjunto de tesis sea consistente. Aceptado esto, el autor se dedica a analizar la ganancia epistémica que la aceptación de causas mentales no físicas implicaría; esto

causa, no un conjunto de causas, y en consecuencia 'ser la causa de' no es una propiedad transitiva. Esto es: si A es la causa de B y B es la causa de C, A no es la causa de C. Dicho de otro modo, sólo se llama 'causa' al penúltimo eslabón de la cadena. Y el principio del cierre causal del sistema físico, concluye, se interpreta normalmente según este sentido de 'causa': todo acontecimiento físico tiene una causa física, y no una mezcla de causas físicas y mentales. Consideramos que la observación es pertinente; sin embargo, no resulta obvio hasta que punto apartarse del uso común debilita el argumento de Lowe; en cualquier caso, no aceptar la formulación del principio de clausura causal impediría desarrollar el argumento en favor del dualismo, por lo que debemos aceptarlo como punto de partida.

es, una vez que hemos aceptado la posibilidad de tales causas. Por esta razón, parece necesario un análisis cuidadoso de los argumentos que nos llevarían a aceptar o rechazar el principio modificado. Podemos observar, de paso, que si este principio modificado se acepta, entonces no habría razones para rechazar, en principio, otros modelos que sostengan otras relaciones entre estados mentales y físicos; por ejemplo, la 'causación descendente' en un esquema que suponga la superveniencia o la realización de lo mental a partir de lo físico. Cabría preguntarse, entonces, si otros modelos de interacción entre sucesos mentales y físicos (por ejemplo, que no apelen solamente a la causalidad como relación de determinación) serían preferibles al propuesto por Lowe.

¿Es razonable afirmar que (4) es más un artículo de fe que un principio plausible para guiar nuestras intuiciones acerca del mundo? Kim (1989b) observa:

> Hay un supuesto ulterior con el cual creo que cualquier fisicalista acordaría, esto es, la 'la clausura causal del mundo físico'. Aproximadamente, dice esto: *cualquier suceso físico que tiene una causa en el momento t tiene una causa física en t*. Este es el supuesto de que si rastreamos los antecedentes causales de un suceso físico, nunca necesitamos ir más allá del dominio físico. Negar este supuesto es aceptar la idea cartesiana de que algunos sucesos físicos necesitan causas no físicas, y si esto es verdad no puede haber en principio una teoría física completa y autosuficiente del dominio físico (p. 280. Cursivas del autor).[141]

Parece innegable que las observaciones de Kim son muy plausibles. Y parecería, también, que la crítica derivada a cualquier posición relativa a la causación mental que pretenda negar este principio también lo es. Sin embargo, diversos autores dudan ya sea de la fuerza del principio, o acerca de la forma correcta de interpretarlo, la cual incide, a su vez, en las consecuencias que se le atribuyen. Convendrá entonces examinar algunos de sus argumentos y analizar la posibilidad de que puedan respaldar la posición de Lowe.

Crane (1995) sostiene que los filósofos rara vez consideran seriamente la posibilidad de negar la completud de la física, entendiendo por tal la tesis de que las causas físicas son completamente suficientes para los efectos físicos; no se requiere ninguna otra causa para producir efectos físicos. El rechazo a negar esta tesis se

debe, en su opinión, a la errónea creencia de que tal negación los comprometería con el dualismo cartesiano. Esta creencia es errónea debido a que ignora la posición según la cual los efectos físicos pueden tener muchos tipos diferentes de causa, ninguno de los cuales tiene los rasgos que Descartes atribuyó a lo mental, y algunos de los cuales no son físicos. Cualquiera sea el significado de 'son suficientes' en el principio de completud de lo físico, este significado debe ser al menos 'suficiente en las circunstancias'. Esto apunta a la cuestión familiar de que raspar un fósforo (se dice) es suficiente para encender un fuego sólo ante la presencia de oxígeno, material inflamable, y así siguiendo. De manera similar, observa, la presencia de oxígeno es suficiente para encender fuego en la presencia del raspado del fósforo y de material inflamable, de manera que es también una causa del encendido del fósforo. Del mismo modo, prosigue, alguien arroja un ladrillo hacia una ventana porque desea hacerlo, y la rompe. Puede estarse de acuerdo con el fisicalista, sostiene, que, *en las circunstancias*, los estados cerebrales de la persona son suficientes para que sus músculos se muevan, para que el ladrillo vuele por el aire, y para que eventualmente rompa el vidrio. Pero las circunstancias también incluyen las creencias y deseos de la persona, por lo que ellas también serán suficientes, dadas las otras circunstancias, para la rotura de la ventana. En cualquier sentido plausible en el cual las causas físicas son suficientes para sus efectos, las causas mentales también lo son. Rechazar la completud de lo físico, por lo tanto, no es rechazar la afirmación de que las causas físicas son suficientes para todos los efectos físicos; es rechazar la afirmación de que *sólo* las causas físicas son suficientes para los efectos físicos. Los efectos físicos, entonces, pueden tener muchas causas: algunas de éstas serán físicas y algunas serán mentales. Según Crane, esta afirmación no está en conflicto con ninguna ley de la física o con ningún principio metodológico perfectamente legítimo; por ejemplo, el principio de que deben buscarse los mecanismos subyacentes de los fenómenos. Debe enfatizarse, advierte, que la aceptación de tal principio no requiere la aceptación del principio de completud de la física. Esta manera de interpretar la tesis de la completud de la física, sostiene, implica una manera de disolver de una manera no fisicalista el problema de la causación mental.

[141] Hay versiones de este principio en términos explicativos: 'en la medida en que un suceso físico (esto es, el cambio en el valor de una propiedad física) pueda ser causalmente explicado, puede ser causalmente

Cierto es que la noción de causa es una noción particularmente controvertida y problemática, por lo que no corresponde legislar acerca de lo que se considera que es *la* manera correcta de concebirla. Sin embargo, notemos un aspecto en el que el enfoque de Crane parece insuficiente. Su posición con respecto a las condiciones suficientes para la ocurrencia de un suceso parece plenamente compatible con el epifenomenismo de lo mental. En efecto, podemos concebir la relación entre el conjunto de condiciones de la siguiente manera: un determinado conjunto de estados y sucesos cerebrales es *causalmente suficiente* para la producción de los movimientos musculares requeridos para arrojar el ladrillo, lo que deriva a su vez en el vuelo de ladrillo y la ulterior rotura de la ventana. Este conjunto de sucesos y estados cerebrales, a su vez, *causa*[142] la ocurrencia de determinados estados y sucesos mentales, como el deseo de arrojar el ladrillo. Pero, en tal caso, los procesos y estados mentales son concomitantes a la existencia de los estados y procesos cerebrales; su existencia como parte de las circunstancias en las cuales se produce la rotura de la ventana no les asegura un rol causal activo en la producción de este suceso. En otros términos, los estados mentales, como las creencias y deseos de la persona, no se encontrarían entre las propiedades *causalmente relevantes* del suceso que identificamos como causa. El contrafáctico 'si m_1 y m_2 (sucesos mentales, el deseo y la creencia) no hubieran acontecido, la rotura de la ventana no hubiera tenido lugar' no garantiza la eficacia causal de tales sucesos.[143]

Aun cuando esta crítica no fuese pertinente, difícilmente los argumentos expuestos en favor de tal interpretación de la completud de lo físico podrían respaldar la posición de Lowe. En el enfoque de éste, como se anticipa en lo expuesto, no se trata de identificar conjuntos de condiciones, tanto físicas como no físicas, que actúen como causas de sucesos físicos o mentales.[144] La ontología presupuesta parece estar constituida por series de sucesos, tanto físicos como no físicos, conectados de manera fundamental por medio de la relación causal; Lowe

---

explicado en esa medida sólo por referencia a otros sucesos físicos' (Loewer, 1995, p. 328).

[142] O, si se prefiere, *determina* vía superveniencia o realización física la producción de tales estados y procesos.

[143] *Cfr.* Kim (1998, capítulo 3) para un análisis de la posibilidad de comprender la causación mental en términos de contrafácticos.

[144] Quizás su enfoque pueda ajustarse al modelo de sucesos como ejemplificación de propiedades (*cfr.* Kim, 1976).

rechaza explícitamente la tesis de que los estados mentales sobrevienen a partir de estados físicos. Por otra parte, aun cuando se acepte que el principio metodológico que prescribe la búsqueda de mecanismos pueda disociarse del fisicalismo, esto no implica, como veremos, que no haya casos concretos en los que la determinación de tales mecanismos no sea exigible.

Baker (1993) señala lo que considera son insuficiencias en la concepción corriente de la clausura causal del mundo físico. Si, como a menudo se sostiene, observa, las propiedades neurofisiológicas son propiedades físicas, entonces no puede interpretarse la tesis de la clausura causal de lo físico de manera tal que afirme que todo suceso neurofisiológico que tiene una causa en *t* tiene una causa neurofisiológica completa en *t*. Esta afirmación sería falsa debido a que las leyes neurofisiológicas contienen cláusulas *ceteris paribus*; en particular, ciertos sucesos a niveles moleculares podrían afectar sucesos neurofisiológicos que están gobernados por leyes neurofisiológicas. Por lo tanto, Baker sostiene que no hay clausura causal al nivel neurofisiológico, ya que un sistema es causalmente cerrado si y sólo si interactúa causalmente *sólo* con otros elementos dentro del mismo sistema. Dado que la clausura causal de la física se aplicaría entonces sólo al nivel más bajo, el de la microfísica (el de las partículas básicas y sus propiedades), propone reformular la tesis de la clausura causal en la siguiente forma: 'Toda ejemplificación de una propiedad microfísica que tiene una causa en *t* tiene una causa microfísica completa en *t*'. Esta reformulación, además de ofrecer reparos en algunos respectos (en particular, parece dudoso que la influencia de los niveles inferiores sobre los niveles superiores sea causal; parecería más razonable fundarla en alguna clase de relación de realización o superveniencia mereológica, modificación que permitiría salvar el principio de clausura causal intranivel) no puede ser utilizada para apoyar la posición de Lowe. En particular, la clase de influencias internivel admitidas por Baker no excede el nivel físico; dicho en otros términos, los sucesos neurofisiológicos, que *también* son sucesos físicos, son afectados por sucesos de niveles inferiores (químicos o aun físicos propiamente dichos), que son sucesos físicos. No hay espacio, en esta reformulación, para acciones causales de supuestos sucesos no físicos sobre sucesos físicos.

Burge (1993), en su análisis de los argumentos en favor de la exclusión causal-explicativa, rechaza que la tesis de la clausura causal tenga la fuerza que usualmente se le atribuye.[145] El razonamiento usual sobre esta cuestión, en su opinión, es el siguiente: todo efecto físico es causado por estados o sucesos físicos previos de acuerdo con leyes físicas aproximadamente deterministas. Las causas mentales originan movimientos físicos en nuestros cuerpos. Si esta clase de causación no consistiera en procesos físicos, existirían desviaciones de los patrones físicos descriptos por las leyes. Sin embargo, no hay razones para pensar que esto ocurra: los estados previos físicos son suficientes para la ocurrencia de los efectos. Y si apelar a la causación mental supone dudar de la adecuación de las formas usuales de explicación física, tal apelación debería ser rechazada.

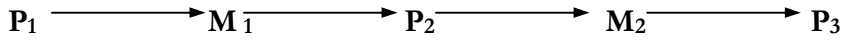
Burge considera que este argumento tiene cierta fuerza, pero no tanta como se le atribuye. Pensar que una causa mental debe interferir o alterar procesos en los sistemas físicos depende de pensar en las causas mentales bajo un modelo físico. La idea, según el, consiste en que la causa debe transferir una fracción de energía o ejecutar una fuerza sobre el efecto. En tal modelo, la causa mental desviaría el efecto físico del curso al que lo conduciría la causa física aislada. Sin embargo, el hecho de que este modelo físico de la causación mental sea adecuado es justamente lo que, en su opinión, está en discusión. Su posición, como hemos visto, es que hay varias formas en las cuales las causas mentales 'hacen una diferencia' [*'make a difference'*] sin entrar en conflicto con explicaciones físicas. En particular, esas diferencias son especificadas por explicaciones psicológicas y por los enunciados contrafácticos asociados a tales explicaciones.

El rechazo al principio de clausura causal no se produce entonces, en la posición de Burge, por un replanteo profundo de las relaciones entre lo mental y lo físico, sino por una exigencia de relegar a la reflexión metafísica y privilegiar a la práctica explicativa. Observemos ahora que las posiciones de Lowe y de Burge son muy diferentes en lo que respecta a esta cuestión. Mientras que Burge rechaza el iniciar la reflexión acerca de la eficacia causal de lo mental a partir de la metafísica, en particular la metafísica materialista, Lowe acepta situar el análisis en el plano

[145] *Supra*, p. 132.

ontológico. Por lo tanto, debería explicar la interacción causal entre sucesos físicos y no físicos. Al respecto, parecen pertinentes dos observaciones.

En primer lugar, conviene observar que habría razones más fuertes que la completud y autosuficiencia de la teoría física para admitir la necesidad de la clausura (como aparece en la definición de Kim); en particular, la incompatibilidad de su negación con la aceptación de principios fundamentales de nuestro conocimiento del mundo físico. No obstante, como veremos en detalle más adelante, Lowe no parece negar la posibilidad de que cualquier suceso físico tenga una causa física; no sostiene, al parecer, que una cadena causal pueda estar constituida de la siguiente forma:



siendo P_n sucesos físicos y M_n sucesos mentales. Todo suceso físico P tendría, desde su punto de vista, una causa física P', lo cual no implicaría la violación del principio de clausura causal. Esto no impediría, sin embargo, que *además* el suceso P tuviera también una causa mental M, sin que se tratara de un caso de sobredeterminación.

Sin embargo, parece haber razones adicionales, y quizás más fuertes que la no completud de la teoría física, para rechazar (3). Este principio, en la interpretación que sostiene Lowe, admitiendo la transitividad de la causación, presenta un panorama aproximadamente así: el mundo es un sistema causal, cuyo origen último es el *big bang* (por lo cual, en última instancia, todo suceso tiene una causa primigenia física), pero en cuya red causal hay nodos tanto físicos cuanto no físicos. Pero esta imagen se enfrenta con un problema serio. La causación de un suceso físico por un suceso no físico parece violar el principio de conservación de la energía (dificultad que, como hemos observado, aqueja a las diversas formas de interaccionismo). Resulta plausible pensar que el hecho de que un estado físico pueda ser causado por un estado no físico implica una violación de este principio, aun cuando no se hable de 'sustancias pensantes' cartesianas. Un estado físico (neurofisiológico) del cerebro involucra cambios fisico-químicos y eléctricos que involucran transferencias de energía, energía que se crearía en el caso de una acción mente-cuerpo, y se destruiría

en el caso de una acción cuerpo-mente. Si esto fuese así, el panorama que se observaría sería el de una creación y destrucción continua de energía, en la medida en que en las cadenas causales se interpongan eslabones no físicos (mentales). Parecería que (3), en la interpretación de Lowe, no es una versión debilitada de (4), sino que implica su negación; esto es, (4) prohíbe tal clase de interacciones, mientras que (3) las permite.

Parece plausible afirmar que la aceptación o el rechazo de la tesis de la clausura causal debe ser una cuestión de 'todo o nada'; dicho en otros términos, y si se nos permite la analogía, podría decirse que la aceptación o rechazo de la clausura causal es similar a lo que representa, para la posición epistemológica de Popper, el rechazo de la inducción; lo que hace Lowe es similar a la aceptación de un 'soplo' de inducción para Popper (y, como ha observado W. H. Newton-Smith sobre esta cuestión, no se trata de un soplo sino de 'una tormenta desatada'). No se puede aceptar una violación, aunque sea parcial, de la clausura, que es lo que en última instancia sostiene Lowe (una red causal en la cual algunos eslabones son no físicos, pero en cuyo origen existe un único suceso físico), y pretender continuar dentro del marco proporcionado por las ciencias físicas de nuestra época.

En segundo lugar, parecería correcto exigir que la postulación de la interacción entre estados mentales y estados físicos vaya acompañada de argumentos positivos en el plano ontológico; en especial, una explicación de los mecanismos que actuarían en tal interacción. Lowe no intenta explicar de que manera un estado físico puede interactuar con un estado no físico sino hasta el final del artículo; su argumento principal consiste en la apelación a supuestas ventajas explicativas que la aceptación de (3) implicaría. Esto es, se nos pide que abandonemos la formulación clásica de un principio que parece estar sólidamente basado en la ciencia básica de nuestra época, en virtud de la supuesta ganancia explicativa que de este modo obtendríamos. Pospondremos entonces el examen de la explicación de la causación entre sucesos mentales y sucesos físicos y, en lo que sigue, analizaremos si la presunta ganancia explicativa es tal; sin embargo, puede presumirse que difícilmente un beneficio explicativo, por notable que sea, logre compensar semejante pérdida de plausibilidad ontológica.

4. Causas mentales y fuerza explicativa

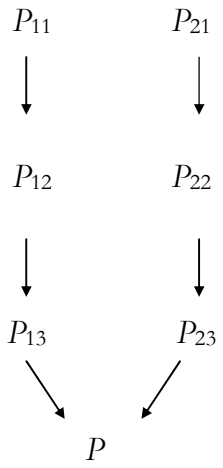
Lowe admite que, aun aceptando que la causación de un estado físico por otro no físico (mental) sea una genuina posibilidad y que las tesis (1), (2) y (3) no son lógicamente inconsistentes, puede pensarse que la sugerencia de que así es como realmente son las cosas es una idea extravagante, que viola principios de parsimonia o simplicidad en temas metafísicos. Por el contrario, sostiene, puede mostrarse que la invocación a los estados mentales, tal como han sido caracterizados, tiene el potencial de fortalecer las explicaciones de determinados sucesos físicos. Esto se debe, prosigue, a que las causas no físicas pueden tornar *no coincidentes* ciertos sucesos físicos, los cuales parecerían ocurrir por mera coincidencia desde la perspectiva de la causación puramente física.

Un suceso ocurre por *coincidencia*, o *coincidentemente*, cuando dos o más sucesos co-ocurren y conjuntamente causan el suceso, pero las causas de estos sucesos son causalmente independientes, dado que ambos sucesos no comparten ninguna causa común entre sus propias causas. Un ejemplo de esta clase de sucesos es el siguiente: un hombre pasa caminando frente a una casa justo en el momento en que una ráfaga de viento desprende y hace caer una teja, la cual golpea al hombre y lo mata. El caminar el hombre y la caída de la teja, ocurriendo conjuntamente, causan la muerte del hombre, pero no hay una causa común entre ambos sucesos. Por lo tanto, la muerte acaece por coincidencia. Un suceso que acaece, como en este caso, por coincidencia, no es un suceso que carezca de causas; las causas coincidentes del sucesos tienen a su vez causas, las cuales tienen causas, y así siguiendo. Lo que permite rotular a este suceso de coincidente, sostiene Lowe, es que sus causas inmediatas no comparten una causa común, por lo cual sus historias causales son independientes.

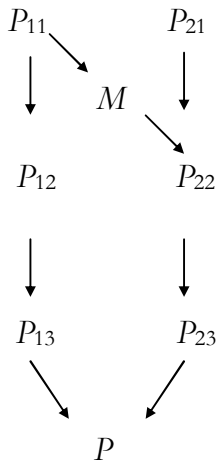
Muy distinta es la situación en la cual los sucesos comparten una causa común. Modificando el ejemplo anterior, si al caminar el hombre hacia la casa hubiera tirado inadvertidamente de un cable atado a la teja, causando su caída justo en el momento en que pasaba por debajo, entonces su caminar frente a la casa y la caída tendrían una causa común, por lo que la muerte del hombre no habría sido coincidente. En el caso del suceso que no ocurre por coincidencia, a diferencia del

que ocurre por coincidencia, el suceso físico del hombre acercándose hacia la casa está relacionado, por medio de la causación física (el movimiento del cable) al suceso físico de la caída de la teja. En este caso, los sucesos en cuestión, tanto en el caso coincidente como en el no coincidente, son puramente físicos. Sin embargo, podría existir una situación en la cual un suceso físico tiene entre sus causas físicas inmediatas una cadena causal que involucre un suceso no físico que vincule ambas historias causales.

Lowe nos pide que consideremos el siguiente diagrama, que describe dos mundos distintos:



Mundo 1



Mundo 2

En este diagrama, las letras representan sucesos particulares, mientras que las flechas significan que el suceso representado por la letra superior es una causa del suceso representado por la letra inferior. Ambos mundos son idénticos en lo que respecta a la ocurrencia del suceso P en la región espacio temporal que nos interesa, y en que los sucesos mantienen las mismas relaciones de causación puramente física unos con otros (esto es, tomando dos sucesos cualesquiera de un mundo, se observará que mantienen las mismas relaciones causales que los sucesos equivalentes del otro mundo). Pero aun cuando en ambos mundos tiene lugar la ocurrencia de P, en el mundo 1 esta ocurrencia es puramente coincidente, mientras que en el mundo 2 no lo es. Si bien P tiene en ambos mundos la misma historia puramente física, en el mundo 2 sus causas inmediatas (P_{13} y P_{23}) comparten una causa común que las

conecta mediante una cadena causal no física, tornando la ocurrencia de P en no coincidente.

La posibilidad de existencia de tales mundos, observa Lowe, significa que aun cuando puedan identificarse todas las causas físicas de determinado suceso físico en una región del espacio-tiempo, esto no excluye la posibilidad de que exista un suceso no físico (mental), como una creencia o un deseo, que pueda servir para explicar por qué el suceso físico es no coincidente, en una forma que la historia causal puramente física no pueda. En sus propias palabras: 'postular ciertas causas no físicas (mentales) de un suceso físico, además de las causas físicas que ya hemos descubierto, puede servir para un propósito explicativo que no puede ser satisfecho apelando sólo a las causas físicas' (p. 234).

La preocupación principal de Lowe al exponer esta posibilidad es mostrar que es perfectamente concebible el hecho de que al investigar un movimiento corporal deberíamos descubrir que todos los sucesos físicos involucrados en su ocurrencia no hacen sino mostrar que el movimiento es puramente coincidente. En este caso su posición es que podemos plausiblemente invocar una causa no física (mental) que torne al movimiento no coincidente, sin negar nada acerca de lo que hasta este momento podamos haber descubierto acerca de las causas puramente físicas.

5. Cómo explicar la causación mental: la direccionalidad de la causación intencional

Hasta aquí, ninguno de los argumentos expuestos por Lowe ha estado destinado a explicar la manera en que un estado o suceso mental puede causar un estado o suceso físico, así como la causación recíproca. Interesado especialmente en la causación intencional (la forma en que un suceso mental intencional puede causar otro suceso *en virtud de ser* un estado intencional), Lowe observa que un rasgo distintivo de ésta es que lo que es causado por tales estados está íntimamente relacionado con el contenido intencional de esos estados. Esta característica hace que las causas intencionales de sucesos físicos siempre estén 'dirigidas' a la ocurrencia de tales sucesos; no parece observarse nada parecido cuando se trata de

la causación de sucesos físicos por otros puramente físicos: tales procesos puramente físicos parecen ser, observa Lowe, 'no dirigidos' o 'ciegos'. Sin embargo, en la acción voluntaria normal un estado intencional del agente es 'dirigido' hacia la realización de un suceso de cierta clase, no hacia el suceso *particular* del cual es la causa. Si nuestro brazo se levanta como resultado de un intento exitoso de elevarlo, lo que es explicado causalmente a través de nuestro deseo no es meramente la ocurrencia de ese caso particular de elevación del brazo, sino la producción de un estado general de cosas en el cual se encuadra un suceso de esa clase que ocurre durante un cierto lapso.

En opinión de Lowe, existe una conexión entre este rasgo de la explicación causal intencional y el rol de las causas mentales en tornar algunos de sus efectos físicos en no coincidentes. Como ha sostenido, lo que hace que un suceso sea no coincidente es un hecho en su historia causal; esto es, el hecho de que sus causas inmediatas no tienen historias causales independientes. En sus propios términos

> Mi nueva sugerencia, ahora, es que este rasgo del rol causal de los estados mentales está íntimamente relacionado con el modo en el cual sirven para proveer explicaciones causales de ciertos estados físicos *generales* de cosas y no meramente de sucesos físicos particulares. Conectando causalmente lo que de otra manera serían cadenas independientes de causación física, sugiero, una causa mental puede volver el efecto común de estas cadenas no coincidente y al hacerlo explicar por qué un suceso de *esa clase ocurrió*, no meramente por qué *ese suceso particular* ocurrió (p. 236. Cursivas del autor).

Esta conexión causal entre sucesos mentales intencionales y sucesos físicos es lo que hace, interpretamos, que el suceso físico no esté sobredeterminado: mientras que la causa física causa que el suceso físico sea ese suceso particular y no otro, la causa mental intencional causa que ese suceso sea de una clase determinada.

Sin embargo, ¿responde este argumento a la pregunta de *cómo causa* un estado mental a un estado físico? En particular, ¿responde a la pregunta de *cómo causa* un estado intencional que un suceso pertenezca a una clase determinada de sucesos? Consideramos que no. Tal como expusimos en el apartado 3., esta respuesta parece requerir de la postulación del mecanismo que vincula un suceso con otro. Podemos tener una explicación de cómo un suceso físico causa otro, en particular de cómo un suceso neurofisiológico en el cerebro causa otro suceso de la misma clase; en otros

términos, podemos *acceder al mecanismo a través del cual* tiene lugar la interacción causal entre sucesos de esta misma clase.[146] Lowe no presenta ninguna explicación de un mecanismo análogo para la interacción entre estados físicos y estados no físicos (mentales). En ausencia de una explicación de este mecanismo no parece que podamos tener acceso epistémico a la interacción causal entre sucesos de clases distintas, como sucesos físicos y sucesos no físicos (mentales). Esta carencia, como hemos observado, parece seria en el enfoque de Lowe, ya que no rechaza fundar la explicación psicológica sobre bases metafísicas.

Por otra parte, no debemos olvidar a los estados mentales no intencionales, pese a que Lowe manifiesta no estar interesado en ellos. Dado que estos estados carecen de contenido, no puede explicarse su presunta acción sobre sucesos físicos por medio de la causación de una clase de suceso. Sin embargo, cualquiera que pretenda mantener la causación mental debe proporcionar una explicación de cómo los estados no intencionales causan sucesos físicos. Por esta razón, en el mejor de los casos la posición de Lowe constituiría una respuesta al problema de la causación intencional, pero no al de la causación mental.[147]

En síntesis, Lowe sostiene que, en ausencia de causas mentales, todo lo que tendríamos serían sucesos coincidentes y explicaciones puramente físicas de sucesos únicos, y no de clases de sucesos, y que nuestras explicaciones se ven fortalecidas si admitimos la causación mental. Ahora bien, ¿es este argumento suficiente para aceptar causas no físicas de sucesos físicos, con las consecuencias ya mencionadas? La respuesta que parece más plausible es la negativa.

Lowe concluye la exposición de su propuesta con una defensa comparativa de su versión del dualismo frente a las versiones más estrictas del fisicalismo:

> [P]*odemos* adoptar una versión de 'dualismo' (por falta de una palabra mejor) que preserve un principio central del fisicalismo, esto es, la afirmación (3): *todo suceso físico tiene un conjunto de causas físicas que son colectivamente causalmente suficientes para la ocurrencia de ese suceso.* Si es sólo una preocupación concerniente a que esta afirmación es

[146] No estamos sugiriendo que dispongamos de todo este conocimiento actualmente, pero no parece existir ninguna razón *a priori* que parezca impedirnos acceder a él.

[147] No hemos considerado, por exceder los objetivos de este capítulo, las objeciones a la eficacia causal de los estados intencionales basadas en las concepciones externalistas del contenido, que constituyen una clase adicional de argumentos en contra de la eficacia causal de los estados mentales de esa clase (*cfr*. Kim, 1998, capítulo 2). Sin embargo, parece razonable pensar que el enfoque de Lowe debería presentar una respuesta a este tipo de argumentos.

> denegada por el dualismo lo que persuade a algunos filósofos a rechazar el dualismo en favor del fisicalismo, entonces espero haber mostrado por qué esta preocupación es bastante errónea. Si, por otra parte, su fisicalismo está motivado por la fe en la afirmación (4) –que *ningún suceso físico tiene una causa no física*- entonces sólo puedo decir que me parece que esa doctrina es un dogma injustificado que los compromete, les guste o no, con el eliminativismo o el epifenomenismo con respecto a lo mental (p. 239. Cursivas del autor).

Sin embargo, la alternativa que propone podría ser considerada plausible en caso de que lograra superar las objeciones referidas a la violación de la clausura causal del mundo físico y a la carencia de una explicación del mecanismo que vincula a lo mental con lo físico. En ausencia de una respuesta a estas objeciones plausibles, podemos concluir que su posición no logra superar las flaquezas tradicionales del dualismo, aun cuando éste sea presentado como 'naturalista'. No cabe ninguna duda de que el materialismo no reduccionista enfrenta severas dificultades que probablemente no pueda superar, pero, por todo lo que sabemos hasta el presente, el dualismo es totalmente inviable.

CAPÍTULO VIII. EL PRINCIPIO DE EXCLUSIÓN EXPLICATIVA

1. Exclusión explicativa y causación mental

En el capítulo I hemos descripto algunos de los argumentos de Kim en favor de la exclusión explicativa en el contexto del problema de la causación mental. Dado que nuestro objetivo principal ha sido describir el problema ontológico de la exclusión para las propiedades mentales, no hemos examinado más en profundidad tales argumentos. Del análisis expuesto, sin embargo, puede inferirse que la exclusión explicativa no se limita al ámbito de la explicación de lo mental, sino que su alcance es sensiblemente mayor, extendiéndose hasta abarcar la competencia explicativa entre teorías rivales.

Por otra parte, el principio ha jugado un rol de importancia variable en el análisis de la causación mental.[148] En algunos de sus primeros escritos sobre el problema de la exclusión causal (1989a, 1990) las consideraciones explicativas jugaron un rol de importancia a la hora de argumentar sobre la impotencia causal de las propiedades mentales. Ya en su (1993), y en particular en el argumento de la superveniencia (1998) la apelación a la explicación ha desaparecido de la argumentación. Este cambio, en nuestra opinión, no ha implicado un debilitamiento de los argumentos sobre la exclusión causal; las consideraciones ontológicas poseen suficiente fuerza probatoria para mostrar la existencia de un problema real para el materialismo no reduccionista.

El principio de exclusión explicativa, ya sea en su versión general o en su aplicación al problema de la causación mental, ha sido objeto de múltiples cuestionamientos. En ocasiones las críticas, si bien se dirigen al rechazo del

[148] Tal fluctuación es menor si se la compara con los cambios que han sufridos sus puntos de vista en relación con la noción de superveniencia, considerada en sus primeros escritos sobre el tema como una posible solución al problema mente-cuerpo y luego, por el contrario, como 'parte del problema'. '[L]a superveniencia no es un *tipo* de relación de dependencia –no es una relación que pueda ser puesta a la par con la dependencia causal, la dependencia reductiva, la dependencia mereológica, la dependencia basada en definibilidad o implicación, y cosas por el estilo-. Más bien, algunas de esas relaciones de dependencia pueden generar la covariación de propiedades requerida y por lo tanto calificar como una relación de superveniencia. La superveniencia, por lo tanto, no es una relación metafísicamente 'profunda'; es sólo una relación 'fenomenológica' entre patrones de covariación de propiedades, patrones que posiblemente son manifestaciones de algunas relaciones de dependencia más profundas' (1998, p. 14).

principio, admiten que la intuición contenida en él es plausible.[149] En este capítulo final quisiéramos plantear una posibilidad no explorada hasta el momento. En caso de que se hallara una respuesta plausible a los argumentos de la exclusión causal, y se lograra presentar una descripción aceptable de la causación mental, ¿podría ocurrir que, de todas formas, el principio de exclusión explicativa pusiera en riesgo las explicaciones basadas en propiedades/sucesos mentales? Dicho en otros términos, que el principio, o una extensión adecuada de él, constituyera el fundamento para la eliminación de los enunciados teóricos explicativos basados en propiedades mentales y su reemplazo por enunciados teóricos explicativos de una neurociencia desarrollada. Si este fuese el caso, el principio de exclusión explicativa cobraría una importancia epistemológica notable, mayor que la que se desprende de su posible aplicación al problema de la causación mental.

A fin de situar con claridad el lugar del principio de exclusión explicativa en el contexto del análisis de la explicación científica propuesto por Kim, describiremos en primer lugar ciertas características básicas que esta autor asigna a la explicación; luego, examinaremos sus limitaciones y las posibilidades alternativas.

1.1. Exclusión explicativa y realismo explicativo

Si bien no ha desarrollado una teorización sistemática sobre la explicación científica, Kim ha examinado algunas de sus características y realizado algunas propuestas en distintos trabajos (especialmente en sus 1987, 1989a y 1994, y comentarios dispersos en varios otros artículos). En estos escritos la posición de Kim se ha caracterizado por una fuerte defensa de una concepción realista de la explicación científica, haciendo énfasis en los aspectos metafísicos de la explicación. Este énfasis le ha conducido a enunciar una serie de críticas a los modelos clásicos de la explicación, en especial a los modelos hempelianos de cobertura legal. Tales críticas han apuntado en especial a la carencia, en la teoría de Hempel, de un análisis de las relaciones de determinación (causalidad, dependencia mereológica, etcétera) que, según Kim, deben ser reflejadas por una explicación científica que pretenda ser

[149] Como hemos visto en el capítulo VI, Marras (1998) ha sostenido que el principio, tal como es propuesto por Kim, conduce a consecuencias inaceptables, por lo que debería ser rechazado; sugiere, sin embargo, que una versión diferente del principio, que evite tales consecuencias, puede ser defendida.

satisfactoria. Así, señala que Hempel sostiene una posición a la cual denomina 'irrealismo' o 'internalismo' explicativo:

> *Irrealismo explicativo* (...) sería la opinión de que la relación de ser un *explanans* para, como aquello que relaciona C y E dentro de nuestro *corpus* epistémico, no está y no necesita estar 'basada' en ninguna relación objetiva entre los sucesos *c* y *e*. Es solamente una cuestión de alguna relación 'interna' entre ítems de conocimiento. Quizás hay relaciones lógicas, conceptuales o epistémicas entre las proposiciones en virtud de las cuales una proposición constituye un *explanans* para otra, y, cuando esto ocurre, podríamos hablar de los sucesos representados como relacionados por una relación explicativa (1987, p. 226. Cursivas del autor).[150]

En contraposición, Kim sostiene que una teoría aceptable de la explicación científica debe caracterizarse por lo que denomina 'realismo explicativo':

> Lo que deseo llamar realismo explicativo adopta la siguiente posición: *C* es un *explanans* para *E en virtud* del hecho de que *c* mantiene con *e* alguna relación objetiva determinada *R*. Llamemos a *R,* sea lo que fuere, una 'relación explicativa' (…) La relación explicativa es una relación objetiva entre sucesos que, podríamos decir, 'fundamenta' la relación *explanans* y constituye su correlato objetivo (*ibid.* Cursivas del autor).

La ausencia de tal requisito en las explicaciones hempelianas, en particular en las nomológico-deductivas, conduce, en opinión de Kim, a ciertas deficiencias explicativas.[151]

En el marco de esta fuerte defensa del realismo explicativo tiene lugar la que posiblemente constituye la principal contribución de Kim a la teoría de la

---

[150] Kim también considera como internalistas a otros teóricos de la explicación, como Friedman y Kitcher.

[151] Kim ha propuesto un aparente contraejemplo al modelo D-N. 'Consideremos primero propiedades disyuntivas en un contexto causal/explicativo. Supongamos que ciertos síntomas médicos pueden ser causados por dos condiciones patológicas bastante distintas. Por ejemplo, tanto el lupus como la artritis reumatoidea causan dolor en las articulaciones (o al menos eso he oído). Supongamos que Mary tiene las articulaciones doloridas, y los exámenes indican que tiene lupus o artritis reumatoidea, pero no sabemos cuál. Consideremos el siguiente argumento "nomológico-deductivo" (una explicación "hempeliana"):

> La artritis reumatoidea causa dolor en las articulaciones. También lo hace el lupus. Mary tiene artritis reumatoidea o lupus. Por lo tanto, Mary tiene dolor en las articulaciones.

¿Tenemos aquí una explicación de por qué Mary está experimentando dolor en sus articulaciones? ¿Sabemos qué causa sus dolores? Creo que hay un sentido perfectamente claro e inteligible en el cual no tenemos aún una explicación: lo que tenemos es una *disyunción de dos explicaciones*, no una *explicación disyuntiva única*. Lo que quiero decir es esto: tenemos dos explicaciones posibles, y sabemos que una o la otra es la correcta, pero no cuál es. Lo que tenemos, sostengo, no es una explicación con una "causa disyuntiva", tener artritis reumatoidea o lupus. No hay tales "enfermedades disyuntivas"' (Kim, 1998, pp. 107-8. Cursivas del autor). Hemos intentado mostrar en otra parte (Fernández Acevedo, 1998) que, en algunos aspectos, las críticas de Kim al modelo N-D no son totalmente justas.

explicación científica: el denominado 'principio de exclusión explicativa'. Este principio, como hemos visto, sostiene básicamente que no puede existir más de una explicación completa e independiente de un mismo suceso. Expresado en otros términos, el principio prescribe que sólo pueden existir dos explicaciones adecuadas para un mismo acontecimiento en el caso en que una de ellas sea incompleta o dependa de la otra.

El origen del principio de exclusión para explicaciones en competencia se sitúa claramente, en el planteo de Kim, en los problemas vinculados con la causación psicofísica. Sin embargo, Kim no se limita a plantear un principio de exclusión para las explicaciones de sucesos mentales, sino que extiende su alcance hasta postular un principio general al que considera 'una restricción general plausible' (1989a, p. 250) y neutral con respecto a las distintas teorías acerca de la explicación científica.

En la perspectiva de Kim, existen argumentos tanto ontológicos como epistemológicos para defender el principio de exclusión. Y si bien su aceptación puede ser defendida, como veremos, sólo a partir de consideraciones gnoseológicas, los argumentos ontológicos, en particular la reflexión sobre la relación causal entre sucesos, aparecen como los de mayor peso argumentativo. Las reflexiones ontológicas que están en la base del problema de la exclusión se vinculan, en el planteo de Kim, con la que él considera una de las dos preguntas fundamentales que deben responder las teorías de la explicación, a la cual denomina 'pregunta metafísica':

> Cuando *G* es una explicación para *E*, ¿en virtud de qué relación entre *g* y *e*, los sucesos representados por *G* y *E* respectivamente, es *G* un *explanans* para *E*? ¿Cuál es la relación objetiva que conecta los sucesos *g* y *e*, que fundamenta la relación explicativa entre sus descripciones *G* y *E*? (1994, p. 56).

Esta pretensión de que las explicaciones deben estar 'basadas' en relaciones objetivas estructurantes conduce a una reflexión acerca de las relaciones de causalidad, superveniencia, dependencia mereológica, etcétera, junto con un análisis de las características de la explicación científica, las que han sido desarrolladas por Kim en diversos trabajos.

Sin embargo, Kim también ha enfatizado que la explicación es una actividad epistemológica, que poseer una explicación es un logro epistemológico, y que las explicaciones que consideremos satisfactorias deben poseer ciertas condiciones de adecuación. Entre otras, Kim considera que cuando buscamos una explicación de *p* típicamente sabemos que *p* es el caso, que la atribución de verdad a la explicación es esencial para la adecuación de ésta, y que no hay necesidad de considerar a la explicación como un argumento o inferencia. La pregunta epistemológica, a su vez, es la siguiente:

> ¿Qué es lo que sabemos –esto es, cuál es exactamente nuestra ganancia epistémica- cuando tenemos una explicación de *p*? (1994, p. 54).

La teorización de Kim sobre la explicación descansa, por lo tanto, tanto en una serie de supuestos ontológicos cuanto en un conjunto de consideraciones epistemológicas, los cuales pueden ser analizados con relativa independencia.

En lo que resta de este capítulo procederemos, en primer lugar, a examinar los fundamentos ontológicos del principio de exclusión explicativa. Con este objetivo en vista, discutiremos la crítica a tales fundamentos propuesta por Sabatés (1996), y propondremos una manera alternativa de concebir la relación entre determinantes y de formular el principio de exclusión explicativa. En segundo lugar examinaremos los argumentos puramente epistemológicos en favor de la exclusión explicativa y sostendremos que no son lo suficientemente fuertes como para justificar el rol que Kim pretende asignar a tal principio.

2. La ontología de la explicación

2.1. La incorrección del PEE: como evitar la exclusión

La aceptación del principio de exclusión explicativa tal como ha sido formulado por Kim requiere de la aceptación de una serie de supuestos ontológicos (relativos a la naturaleza de la causalidad, de la superveniencia, etcétera) y epistemológicos (referentes a la teoría de la explicación científica y a los requisitos de simplicidad, unificación y coherencia entre los ítems de conocimiento, entre otros). La pregunta que surge casi naturalmente a partir de esta descripción es

¿constituyen estos supuestos un conjunto consistente y aceptable? Sabatés (1996) proporciona una respuesta negativa a este interrogante. En su artículo desarrolla una elaborada crítica al planteo del principio de exclusión de Kim, mostrando que la aceptación de ciertos supuestos que parecen correctos conduce a una inconsistencia, y concluyendo la presunta necesidad de renunciar al principio para evitarla. En este apartado examinaremos esta crítica, y sostendremos que si bien es acertada en lo que se refiere a señalar algunas insuficiencias argumentativas del planteo de Kim, pueden explorarse algunas posibilidades alternativas, una de las cuales parece preservar el principio de exclusión en términos similares a los que fue originalmente formulado.

La inconsistencia en el planteo de Kim, de acuerdo con la crítica de Sabatés mencionada, surge a partir de la aceptación simultánea de los siguientes supuestos:

1. Realismo explicativo: las explicaciones están basadas en relaciones estructurantes objetivas que determinan los sucesos del mundo; las consideraciones lógicas o epistémicas, si bien necesarias, no son suficientes para garantizar la 'corrección' o 'verdad' de una explicación.
2. Pluralismo explicativo: existen explicaciones que no son causales, esto es, explicaciones basadas en relaciones de determinación distintas a la causalidad, tales como la relación de dependencia mereológica (o 'superveniencia mereológica'), la identidad, etcétera.
3. Exclusión explicativa: no puede haber más de una explicación completa e independiente de un mismo suceso.

Si a estos supuestos plausibles se les agrega un cuarto supuesto, consistente en afirmar que:

4. Todo suceso tiene una causa (al cual podríamos llamar 'principio de determinismo causal'),

se genera una inconsistencia que, según Sabatés, puede ser resuelta mediante el rechazo del principio de exclusión, pero sin abandonar las intuiciones que lo motivan.

El núcleo de su argumento puede resumirse en lo que sigue. El realismo explicativo supone la existencia de 'relaciones estructurantes objetivas', y puede pensarse que la relación causal es el candidato privilegiado a ocupar el rol de esa relación. La tesis de que la relación causal es la única relación estructurante objetiva

podría denominarse 'causalismo explicativo'. Pero, prosigue Sabatés, se ha argumentado que el causalismo explicativo constituye una posición demasiado estrecha. Un criterio más liberal consiste en sostener el pluralismo explicativo (PE), de acuerdo con el cual existen, además de explicaciones basadas en la relación causal, otras explicaciones fundamentadas en otras relaciones estructurantes en el mundo, tales como la dependencia o superveniencia mereológica,[152] la identidad, y la 'dependencia de Cambridge'. La existencia de la relación de superveniencia mereológica es suficiente, según Sabatés, para que el pluralismo explicativo resulte más plausible que el causalismo explicativo.

El problema se presenta cuando a los supuestos mencionados se añade la creencia plausible de que no existen sucesos no causados. Dado que todo suceso tendría una causa, existiría una explicación causal para cada uno de ellos. Si a esto se suma que el principio de exclusión explicativa prescribe que sólo puede existir una explicación completa e independiente de cada suceso, Kim se vería llevado a admitir que no podrían existir más que explicaciones causales para los sucesos del mundo, con lo cual el pluralismo explicativo resultaría falso.

La solución que Sabatés propone a este problema consiste en rechazar el principio de exclusión explicativa, aunque, sugiere, sin negar las ideas que subyacen al mismo. Señala correctamente, a nuestro modo de ver, que los argumentos de Kim en favor de este principio descansan exclusivamente en consideraciones causales, y que estos argumentos conducen únicamente a una formulación de un principio de exclusión causal y, consecuentemente, a un principio de exclusión para explicaciones causales. El primero, de exclusión causal (EC), sostendría que no pueden existir dos causas completas e independientes de un mismo suceso,[153] mientras que el segundo, al que podríamos denominar de exclusión de explicaciones causales (EEC), afirmaría que no puede haber dos explicaciones *causales* completas e independientes de un mismo suceso. Sabatés advierte que el principio general de exclusión explicativa de

[152] La relación de superveniencia mereológica es la relación por la cual la ejemplificación de una propiedad en un objeto depende de las propiedades y las relaciones ejemplificadas entre sus partes (por ejemplo, la propiedad de ser transparente de un fluido depende de las propiedades y relaciones entre las moléculas del fluido).

[153] Excepto en los casos, más bien raros, de sobredeterminación causal. Kim considera, no obstante, que aún en esos casos puede sostenerse que no podemos tener dos explicaciones completas e independientes del suceso sobredeterminado (*cfr*. Kim, 1989a, p. 252). *Cfr.* al respecto la nota 14.

Kim consiste en una extensión injustificada a partir del principio de exclusión causal, ya que, aun admitiendo que las consideraciones causales son automáticamente relevantes para las explicaciones, sus conclusiones se aplicarían solamente a las explicaciones causales. Dado que el principio de exclusión causal conduce solamente al de exclusión para explicaciones causales, concluye Sabatés, constituye un *non sequitur* el intento de utilizarlo para excluir algo más que explicaciones causales rivales de un mismo acontecimiento.

Por otra parte, a partir de ciertos argumentos particulares en favor de EC (en particular, los argumentos en contra de una sobredeterminación causal global de la realidad) no pueden construirse argumentos análogos para defender el principio general de exclusión explicativa. En particular, dice Sabatés, debe advertirse que quien ofrece una explicación no causal de un suceso *b* no se compromete a la vez con la existencia de una causa sobredeterminante de *b*. En un caso como éste, señala Sabatés, si damos una explicación no causal de la ocurrencia de un suceso *e*, no nos comprometemos con la existencia de una causa sobredeterminante de *e*:

> Supongamos que tenemos una explicación mereológica de un macrosuceso *e* en términos de su base de dependencia *b*. Supongamos que tenemos también una explicación de *e* que involucra su causa *c* (…) La explicación no causal no apela a ninguna relación causal que rivalice con la relación causal entre *c* y *e*, por lo que *b* (la base de dependencia de *e*) no constituye una causa sobredeterminante de *e*. De este modo, *b* no está en posición de excluir causalmente a la causa microfísica *c* y el fisicalismo no se ve amenazado (p. 105).

A esto podría responderse, sostiene, que la sobredeterminación constituye un problema ya que, una vez que se observa un determinante suficiente para un suceso, el otro determinante podría no estar presente y el suceso ocurriría de todas maneras. En este sentido, la base mereológica de un acontecimiento sería un determinante suficiente para la ocurrencia de un suceso, y en cierta forma excluiría, si bien no causalmente, a la causa del acontecimiento. Este principio de exclusión modificado, que podría ser denominado 'principio de exclusión de determinantes', sostendría que no puede haber dos determinantes suficientes para el mismo suceso. Si consideramos únicamente la relación causal y la de superveniencia, esto implicaría la

existencia, por una parte, de sucesos no causados, y por la otra, de sucesos no sobrevinientes.

En algunos de sus artículos el propio Kim parece admitir la posibilidad de que existan sucesos no causados:

> El lugar dominante conferido a la relación causal es evidente en el hecho de que, por ejemplo, la tesis del determinismo universal es a menudo formulada en alguna forma como 'todo suceso tiene una causa'. La suposición implícita en tal formulación es que ser determinado viene a ser lo mismo que ser causado. No obstante, esto requiere ser reconsiderado. Parece haber relaciones de dependencia entre sucesos que no son causales, y, como argumentaré, el determinismo universal puede ser verdadero aun cuando no todo suceso tenga una causa (1974, p. 22).

De manera consistente con este punto de vista, Kim ha considerado a la superveniencia mereológica como una relación estructurante distinta y complementaria de la causalidad, la cual nos proporciona un principio de determinación distinto al aportado por la relación causal

> Pienso que es iluminador ver al determinismo causal y el determinismo mereológico como principios regulativos complementarios con raíz metafísica: el primero nos da un principio de determinación diacrónica y el último un principio de determinación sincrónica. Exactamente tal como el determinismo causal afirma que los estados del mundo están ordenados determinísticamente a lo largo de la coordenada temporal, el determinismo mereológico afirma que dentro de cada sección transversal del mundo hay relaciones de determinación que caracterizan los diferentes componentes del mundo' (Kim, 1978,[154] p. 154, citado en Pérez, 1996).

No obstante, el principio ontológico de exclusión modificado de la forma expuesta (PED) no implica simplemente la admisión de la existencia de sucesos no causados: elimina la posibilidad misma de la causación. Si bien es más fácil de obtener que el principio general de exclusión explicativa, sostiene Sabatés, es altamente implausible. Si se tiene la razonable convicción de que todo macrosuceso depende de sucesos microfísicos, entonces *todo* macrosuceso sería no causado. Por lo tanto, el principio de exclusión de determinantes conduce a resultados inaceptables, por lo que no puede ser adoptado. Y sin él, concluye Sabatés, no existen bases para la exclusión al nivel explicativo.

[154] (1978), 'Supervenience and Nomological Inconmensurables', *American Philosophical Quarterly*, vol. 15.

2.2. Múltiples determinantes y completud explicativa

¿Son los argumentos de Sabatés suficientes para desechar el PEE? Responder a esta pregunta requiere, nos parece, plantear una serie de condiciones. El problema de si el PEE es un principio plausible aun cuando admitamos otras relaciones además de la relación causal sólo puede tener una solución condicional si no se especifican las otras relaciones objetivas que pueden servir de base para una explicación. Por otra parte, el problema no sólo se complica por la existencia de varias relaciones aspirantes a relaciones estructurantes objetivas, sino también por el hecho de que las explicaciones psicológicas no solamente (y quizás, si seguimos en esto a Cummins, ni siquiera principalmente) están dirigidas a explicar sucesos, sino también estados, procesos o disposiciones. La carga de la prueba recae sobre el defensor de la exclusión si desea demostrar que no puede haber más de un determinante (y, consecuentemente, una explicación) para cada suceso; pero le corresponde al oponente de la exclusión si pretende probar que ésta lleva a resultados inaceptables: debe mostrar que hay otras relaciones objetivas de determinación que permiten construir explicaciones no causales de sucesos individuales, y que estas explicaciones no compiten con las explicaciones causales. Trataremos de mostrar que si nos limitamos a dos clases de relaciones objetivas, la causalidad y la dependencia o superveniencia mereológica, parece haber maneras de conciliar las relaciones manteniendo la exclusión explicativa.

Las relaciones de determinación que son a menudo mencionadas como posibles bases ontológicas para la construcción de explicaciones no causales de sucesos individuales son varias: la ya mencionada dependencia mereológica, la superveniencia, la dependencia de Cambridge, la relación determinado-determinable, la identidad, la 'dependencia de acciones' [*agency dependence*] la 'composición de sucesos' [*event composition*] (*cfr.* Kim (1974); Ruben (1990)).[155] Ahora bien, el pluralismo explicativo no requiere de todas esas relaciones; basta con admitir (lo cual parece incontrovertible) que la relación de dependencia mereológica es una

[155] No es necesario aceptar, por otra parte, que toda relación entre hechos deba ser explicativa. Recordemos lo que sostiene Ruben: así como no toda relación causal es explicativa (ya que depende de cómo causas y efectos son descriptos), no toda relación de identidad es explicativa. Si bien la afirmación de identidad temperatura = energía cinética media es explicativa, la afirmación temperatura = temperatura no lo es.

relación que tiene lugar entre estados y propiedades y que puede fundamentar explicaciones bajo un enfoque realista.[156]

Una pista para desarrollar una interpretación que permita conservar el PEE con la modificación señalada es una observación de Kim referente a los dos tipos de explicación distinguidos por Cummins:

> Robert Cummins ha realizado una interesante distinción entre explicación 'por subsunción' (bajo una ley causal) y explicación 'por análisis' (en partes componentes) (...) ¿Podrían darse explicaciones de estos dos tipos de un mismo *explanandum*? De acuerdo con Cummins, no obstante, la explicación subsuntiva explica *cambios* y la explicación analítica explica *propiedades*, por lo que los *explananda* son diferentes. Parece también posible interpretar las dos explicaciones como mutuamente complementarias, pero siendo cada una de ellas una explicación parcial de un *explanandum* único bajo una individuación no detallada de los *explananda* (1989a, p. 258, *n*. 39).

En efecto, Cummins ha observado que muchas teorías científicas no están diseñadas para explicar cambios sino para explicar propiedades. Estas teorías están destinadas a explicar las propiedades de un sistema no en el sentido planteado por las preguntas '¿Por qué S adquirió P?' o '¿Qué causó que S adquiriera P?', sino en el sentido planteado por preguntas tales como '¿en virtud de qué S tiene P?' Mientras que la pregunta característica respondida por una teoría transicional es '¿por qué el sistema S cambia del estado s-1 al estado s-2?', la pregunta característica planteada por una teoría de propiedad [*property theory*] es: '¿para qué tiene la propiedad S el sistema P?' Para responder a esta última pregunta la estrategia típica es construir un análisis de S que pueda explicar la posesión de P por parte de S apelando a las propiedades de los componentes de S y su modo de organización.[157]

Siguiendo esta línea de análisis, un *explanandum* no especificado podría ser '¿por qué es el caso que *p*?', siendo *p* la posesión de una propiedad por parte de un objeto. Bajo esta individuación no precisa serían posibles dos explicaciones no

[156] Por otra parte, un pluralista explicativo 'austero' podría tener razones para centrarse sólo en el análisis de las relaciones de dependencia mereológica y de causalidad. Por ejemplo, es posible pensar que la relación determinado-determinable no sea una relación objetiva que tiene lugar entre propiedades, sino entre predicados; la 'dependencia de acciones' se encuentra limitada al ámbito de las acciones humanas; la superveniencia (al menos si le creemos a Kim en sus últimos escritos) no es una relación 'metafísicamente profunda', sino una relación meramente fenomenológica y dependiente de relaciones más básicas.

[157] Cummins señala que el análisis es 'recursivo', dado que un análisis determinado puede recurrir a propiedades o componentes que a su vez requieran análisis.

competitivas: una, causal, referente al conjunto de condiciones cuya modificación hizo posible la aparición de *p*; otra, por dependencia mereológica, referente a las propiedades de nivel inferior que son su base. En este primer caso, las explicaciones podrían considerarse parciales y no competitivas.[158]

Un caso en el que individualicemos con más precisión el *explanandum* sería muy diferente. Consideremos estos dos *explananda* alternativos al anterior: '¿cómo ha llegado la entidad *x* a adquirir *p*?', y '¿cuáles son las condiciones que permiten la existencia de *p*?' En este caso, será apropiada una explicación causal que especifique las condiciones que han sufrido cambios, posibilitando la adquisición de *p* por parte de *e*. Debido a que del grado de especificación del *explanandum* dependerá la cantidad y el tipo de explicaciones posibles, una explicación causal completa podrá ser, en algunos casos, equivalente a explicación completa *simpliciter*, mientras que en otros será sólo una parte de una explicación mayor que la incluya. Lo mismo podría afirmarse, *mutatis mutandis*, de una explicación por dependencia mereológica.

Cuando explicamos un estado, entonces, nos bastará con una explicación que haga referencia a las propiedades y entidades de nivel inferior de las cuales depende; no estaremos explicando *cambios* que requieran hacer referencia a modificaciones en las condiciones antecedentes. Lo anterior no implica afirmar que la relación causal que sirve de fundamento a una explicación causal de un suceso no pueda, eventualmente, ser objeto de explicación. Si aceptamos, como lo hemos hecho, que todo macrosuceso depende mereológicamente de sucesos de nivel inferior, nos comprometemos con la afirmación de que no puede haber cambios en los estados de nivel superior que no estén basados en cambios de estados de nivel inferior. Una manera de conciliar ambos tipos de determinaciones (causal y por dependencia mereológica) es por medio de la noción de 'causación superveniente'. Las nociones de 'causación epifenoménica' y 'causación superveniente' han sido desarrolladas por Kim en sus (1984a) y (1984b), con un interés especial en su aplicación para el caso de la causación psicofísica. Las examinaremos a continuación.

[158] Hemos visto en el capítulo VI que Thomasson sostiene que los dos determinantes de la producción de un suceso no compiten entre sí, y tampoco lo hacen las explicaciones respectivas. Sin embargo, Thomasson no distingue entre la explicación de sucesos y la explicación de propiedades (estados), por lo que su caracterización parece incompleta.

Por 'macrocausación' Kim entiende relaciones causales que involucran macrosucesos y (macro)estados, entendiendo éstos como la ejemplificación de una macropropiedad por un objeto en un momento determinado.[159] Kim rechaza la presunción de que la macrocausación sea un rasgo irreductible del mundo; por el contrario, sostiene que un precepto metodológico fundamental de la ciencia física moderna consiste en tratar de formular teorías microestructurales de los objetos y sus propiedades, esto es, tratar de entender su conducta en términos de las propiedades y relaciones que caracterizan a sus microconstituyentes. La suposición filosófica que subyace a esta estrategia de investigación es la creencia de que las macropropiedades están determinadas por, o sobrevienen a partir de, micropropiedades. La noción de suceso superveniente es explicada a partir de la descripción anterior de una propiedad superveniente. Un suceso *x* que posee la propiedad F sobreviene a partir de un suceso *x* que posee la propiedad G sólo en el caso en que *x* posea la propiedad G y G sea la base de superveniencia de F.

De acuerdo con esta descripción, es plausible pensar que podemos disponer de dos explicaciones de la relación existente entre dos macropropiedades F y G. Una de estas explicaciones será una explicación macrocausal que vinculará a F y G. Esta explicación no hará referencia a ningún otro factor que no sea la macroley causal que vincula a F y G y, posiblemente, a determinadas condiciones iniciales. Una segunda explicación deberá hacer referencia a la microley causal que vincula a las micropropiedades a través de las cuales sobrevienen F y G; la primera explicación será, en términos de Kim, reductible a la segunda. Un ejemplo de esta clase de relación, al que Kim denomina standard, es el de la explicación del aumento de la temperatura de un gas por efecto del aumento de la presión. Esta relación macrocausal, sostiene, queda subsumida bajo una macroley, la ley de los gases, la cual es a su vez microrreducida por la teoría cinética de los gases.[160]

[159] Como Kim hace notar en su (1984b), la distinción 'micro-macro' es relativa '…la temperatura es macro en relación con el movimiento molecular; las propiedades de las moléculas son macro en relación con las propiedades y relaciones que caracterizan a los átomos y a las partículas más básicas, y así siguiendo' (p. 95). Si este es el caso, tendríamos que admitir que los únicos sucesos no sobrevinientes serían los que tienen lugar en el nivel de los componentes más básicos de la realidad –quarks, o cuales fueren-; únicamente a este nivel podría hablarse, con propiedad, de sucesos 'micro'.

[160] Es interesante observar que, para Hempel, la ley que relaciona las variables que determinan el comportamiento de un gas no es causal, sino que se trata de un caso que denomina *leyes de coexistencia* (1965, p. 347).

El modelo de reducción general para una relación causal entre dos macrosucesos, el suceso *x* que posee la propiedad F y el suceso *y* que posee la propiedad G (siendo F y G macropropiedades), requiere de la existencia de una micropropiedad m(F) que es poseída por *x* y sobre la cual sobreviene F, y la existencia de una micropropiedad m(G) que es poseída por *y* sobre la cual sobreviene G; existe además entre *x* que posee m(F) e *y* que posee m(G) una relación causal que las vincula. Para Kim, cualquier relación causal que satisfaga este patrón es un caso de causación superveniente.

Por otra parte, sería erróneo suponer que, en la perspectiva de Kim, el hecho de que la mayor parte de las relaciones causales que observamos sean casos de causación superveniente implica que tales fenómenos sean de alguna manera menos 'reales' que los microsucesos a partir de los cuales sobrevienen:

> Por otra parte, la relación causal entre el aumento de la temperatura y el incremento de la presión de los gases no es menos 'real' por ser microrreducible. Tomar la microrreducibilidad como una impugnación de la realidad de aquello que es reducido haría que todo nuestro mundo observable fuese irreal (1984, p. 102).

En el caso de la causación superveniente, por lo tanto, 'no tenemos (...) dos explicaciones causales *independientes* del mismo acontecimiento. Ambas explicaciones pueden coexistir porque una de ellas depende de la otra, de manera reductiva o por superveniencia' (1989a, p. 251). La reducibilidad de la macrocausación a la microcausación resulta un rasgo inaceptable de la causación superveniente para algunos autores, que si bien quieren mantener las relaciones causales de nivel superior dependientes de las de nivel inferior, pretenden mantener a la vez no sólo su realidad sino también su irreductibilidad. Sin embargo, dadas las múltiples posibilidades que parece ofrecer el concepto de superveniencia, no parece inevitable la admisión de que tal relación de dependencia implique la reducibilidad.

Cualquiera de las dos estrategias posibles que hemos descripto (la división del *explanandum* y la idea de dos explicaciones complementarias de un único *explanandum* poco determinado) parece salvar al principio de exclusión explicativa. En el primer caso, ya no hay dos explicaciones en competencia de un mismo *explanandum*; en el segundo, hay dos explicaciones parciales y conjuntamente completas. Esta

posibilidad obliga, como hemos anticipado, a formular el principio en términos de la imposibilidad de dos explicaciones completas e independientes de un mismo *explanandum*, y no de un mismo suceso. La conclusión que admitiremos estará formulada, entonces, en términos condicionales: si aceptamos un pluralismo explicativo mínimo (esto es, uno que sólo acepte relaciones causales y por dependencia mereológica), entonces podemos aceptar la exclusión explicativa para un mismo *explanandum*. No podemos afirmar, además, que no haya maneras similares de coordinar la causalidad con las otras relaciones estructurantes de manera tal que las explicaciones respectivas no compartan los *explananda*, o que sean explicaciones parciales de un *explanandum* caracterizado con escasa precisión.

3. El principio de exclusión explicativa: su justificación epistemológica

3.1. Internalismo explicativo y exclusión explicativa

Como hemos visto, Kim pretende que el principio de exclusión explicativa constituye una 'restricción general plausible' para las explicaciones en general. También pretende que el principio constituye una condición aceptable sea cual fuere la teoría de la explicación que se acepte:

> Un examen detallado de la exclusión explicativa inevitablemente se extenderá hasta abarcar el antiguo debate acerca de la naturaleza de la explicación, un tema sobre el que no existe nada parecido al consenso en la actualidad (…) sin embargo, espero que la discusión tenga éxito en mostrar que cualquiera sea el modelo explicativo que se acepte, a menos que se adopte una concepción completamente ficcionalista o instrumentalista de la explicación, el principio de exclusión explicativa es una restricción general plausible (1989a, p. 250).

De acuerdo con Kim, cuando buscamos una explicación nos encontramos en una situación de predicamento epistémico, predicamento que se repite cuando disponemos de dos (o más) explicaciones para un mismo hecho. Sugiere, basándose en la teoría de la explicación desarrollada por M. Friedman y posteriormente por P. Kitcher, que esta situación se produce debido a que las explicaciones múltiples de un mismo *explanandum* son contraproducentes en relación con el *desideratum* de simplicidad y unificación entre los ítems de nuestro conocimiento. Por lo tanto, si la simplicidad y la unificación de la teoría constituyen la meta de la búsqueda de

explicaciones, la presencia de múltiples explicaciones de un mismo hecho atenta contra este objetivo, a menos que pueda determinarse que las distintas premisas explicativas se hallan apropiadamente interconectadas.

Si en su versión causal el principio de exclusión sostiene que no puede haber *más de una causa* para un mismo hecho o suceso (excepto en los raros casos de sobredeterminación causal), en su versión explicativa el principio postula que no puede haber dos (o más) explicaciones *completas e independientes* de un mismo fenómeno. Pero Kim considera conveniente distinguir dos versiones diferentes del principio de exclusión explicativa: el principio *metafísico* de exclusión sostiene que dos explicaciones diferentes de un mismo *explanandum* 'pueden ser ambas explicaciones correctas sólo si al menos alguna de las dos es incompleta o una es dependiente de la otra' (1989a, p. 257); el principio *epistemológico* sostiene que nadie puede aceptar dos explicaciones 'a menos que tenga una explicación apropiada de cómo están relacionadas entre sí' (*ibid.*, p. 257). Kim agrega que, en su versión metafísica, el principio puede aceptarse sólo si a la vez se acepta alguna forma de *realismo explicativo*, ya que las nociones de incompletud o dependencia implican alguna relación objetiva en el mundo. Pero, observa, en su versión epistemológica el principio es aceptable aun cuando se renuncie a la idea de que existe alguna relación explicativa objetiva en el mundo, ya que puede encontrarse perturbador y disonante el hecho de tener que tratar con dos (o más) explicaciones independientes del mismo fenómeno. No es necesario que los respectivos *explanans* sean lógicamente contradictorios, e incluso pueden tenerse suficientes garantías en favor de la verdad de cada conjunto de premisas. Sin embargo, aceptar ambos podría inducir a algún tipo de incoherencia en nuestro sistema de creencias. De este modo, '*demasiadas explicaciones pueden ser una fuente de incoherencia en vez de incrementar la coherencia*' (*ibid.*, p. 258. Cursivas del autor).

Los principios de simplicidad y unificación de nuestro sistema de creencias y la teoría coherentista de la justificación son, entonces, los argumentos a los que el irrealista explicativo puede apelar para justificar su aceptación del principio de exclusión explicativa. Si la teoría coherentista de la justificación es uno de los fundamentos para la aceptación del principio epistemológico, podemos

preguntarnos en primer lugar en qué medida esta teoría y el irrealismo explicativo se combinan adecuadamente entre sí.

Recordemos que Kim sostiene que el requisito de verdad de las premisas explicativas, condición necesaria para una concepción realista de la explicación, no es, sin embargo, una condición suficiente. Un irrealista explicativo, rechazando la idea de que la explicación requiere de la existencia de una *relación objetiva* entre los sucesos involucrados, puede igualmente hablar de verdad o corrección de la explicación, teniendo en cuenta la verdad de sus enunciados componentes.[161]

Las tesis características de las teorías coherentistas de la justificación, según Haack (1995), son las que afirman que

> [L]a justificación es exclusivamente un asunto de relaciones entre creencias, y que es esta coherencia de las creencias dentro de un conjunto lo que justifica a las creencias miembros de ese conjunto. Diré que una teoría califica como coherentista si suscribe la siguiente tesis:
> (CH) Una creencia es justificada si y sólo si pertenece a un conjunto coherente de creencias (p. 17). [162]

Una posición que sostenga simultáneamente las tesis del irrealismo explicativo con las de una teoría coherentista de la justificación debe ser evaluada como internamente consistente. Si lo que buscamos cuando intentamos proporcionar una explicación de un suceso es que sus proposiciones componentes tengan cierto tipo de relación, lógica o epistémica, resultará deseable la coherencia general del sistema de proposiciones del cual forma parte esa explicación. Por otra parte, como Haack señala, la 'coherencia explicativa', ha sido uno de los criterios

[161] Nótese que el internalismo explicativo, tal como ha sido caracterizado por Kim, no afirma que una explicación que sostenga la existencia de un correlato objetivo para su *explanans* es una explicación inadecuada o incorrecta; sólo sostiene que la existencia de este correlato no es *necesaria* para la corrección de una explicación. Además, el internalismo no necesita ser adoptado de una manera global: es plausible pensar que un internalista puede considerar adecuada a una explicación aunque ésta postule la existencia de un vínculo objetivo entre los sucesos que se ven referidos en el *explanans*, pero puede a la vez admitir como correcta a otra explicación de un suceso distinto que no lo haga.

[162] Haack advierte que hay lugar para ciertas variaciones de criterio acerca de que características debe tener el conjunto para ser considerado coherente. Algunas de las variantes discutidas son la consistencia, la 'comprensividad', y la coherencia explicativa. También hay variantes con respecto al *status* de las diversas creencias que forman parte del conjunto: mientras que algunos sostienen que todas las creencias que pertenecen al conjunto se hallan en pie de igualdad con respecto a su justificación, otros afirman que ciertas creencias tienen un lugar privilegiado dentro del conjunto. Todas estas distinciones, no obstante, son refinamientos de la tesis principal, y no necesitamos distraernos en su análisis. Por otra parte, dado que Kim no indica cual forma de coherentismo considera aceptable, no hace falta profundizar en la cuestión, sino atenernos a la formulación más general de esta posición.

propuestos para esclarecer el significado de la noción 'conjunto coherente de creencias'. No es en los problemas de la compatibilidad entre la explicación como unificación y la teoría coherentista, al parecer, donde pueden surgir dificultades con este argumento de Kim.

Sabatés (1996) ha señalado que la defensa del principio de exclusión explicativa basándose en consideraciones que denomina de 'plausibilidad y elegancia' es insuficiente. Considera, en primer lugar, que no existen fundamentos ontológicos para la exclusión, por lo cual la exclusión explicativa basada en tales consideraciones constituye un *non sequitur*. No obstante, señala, el defensor de la exclusión explicativa podría argumentar que una 'superpoblación' explicativa iría en contra del carácter unificador y simplificador propio de la actividad explicativa. La existencia de distintas explicaciones de un mismo suceso parece ir en una dirección opuesta al principio de simplificación explicativa: explicar lo más que se pueda con el menor número de premisas explicativas. La coexistencia de explicaciones competitivas, entonces, parece constituir una desventaja epistemológica.

Sin embargo, señala Sabatés, esta consideración fracasa en su objetivo de proporcionar una base para la exclusión explicativa. Lo que es considerado una buena explicación depende de la situación epistémica en la cual se encuentran quienes necesitan mejorar su comprensión de un fenómeno. Puede ocurrir que, en ciertas situaciones epistémicas en las cuales contamos con dos explicaciones en competencia, cada una de ellas pueda mejorar nuestra comprensión del *explanandum*; además, es posible que seamos capaces de proporcionar una sola de las explicaciones 'rivales', pero no la restante. Parece claro, concluye Sabatés, que existen *ganancias epistémicas* como resultado de disponer de dos explicaciones del mismo fenómeno, y estas ganancias deben ser contrapuestas a la carencia de simplicidad. Estas consideraciones de Sabatés pueden proporcionar la base para un argumento en contra de la aceptación irrestricta del principio de exclusión explicativa en su versión epistemológica sin admitir el principio en su versión metafísica.

En primer lugar, se debe hacer notar que, aun cuando existieran bases ontológicas para la exclusión explicativa, esto no es algo que debería preocupar a quien desee defender la versión epistemológica del principio de exclusión sin aceptar

la versión metafísica; por ejemplo, a un internalista explicativo. Si, como vimos, para que una explicación sea correcta no hace falta que exista un correlato objetivo para su *explanans*, el único argumento para no aceptar a dos explicaciones acerca del mismo *explanandum* es el argumento de simplicidad y unificación. Por esta razón, la coexistencia de explicaciones rivales es mucho más problemática para un realista que para un internalista, ya que para el primero no sólo estarán presentes las consideraciones de simplicidad y unificación, sino también, y muy especialmente, que bajo la exclusión ontológica *no pueden ser ambas completas e independientes*. Bajo el internalismo explicativo, por el contrario, dos explicaciones de un mismo *explanandum* no necesitan ser evaluadas a partir de ningún requisito de completud, ya que no se postula la existencia de una relación objetiva que les sirva de correlato; por otra parte, ambas podrán ser independientes en la medida que sean epistemológicamente independientes sus respectivos *explanans* (esto es, que entre cada conjunto de enunciados distintos e independientes se establezcan las relaciones lógicas, conceptuales o epistémicas que se requieran para constituir una explicación).

Intentemos ahora determinar en qué circunstancias dos explicaciones independientes de un mismo *explanandum* pueden constituir una mejora en nuestra situación epistémica. Sean *A* y *B* dos explicaciones de un mismo suceso *s*, digamos una conducta de un organismo *O* en una situación *S*. *A* es postulada a partir de una teoría acerca de lo mental que acepta la existencia de creencias, deseos e intenciones; *B* es postulada a partir de una teoría neurológica que sostiene la existencia de mecanismos neurofisiológicos, sin hacer referencia a creencias o deseos. Hay elementos de juicio para pensar que ambas teorías son verdaderas; de hecho, también para creer que los respectivos *explanans* están compuestos por afirmaciones verdaderas. Pero, supongamos, bajo una interpretación internalista de ninguna de las dos explicaciones se pretende que tengan un correlato objetivo para sus *explanans* (excepto en lo que se refiere a la existencia de las entidades y propiedades a las que se hace referencia en la explicación). Hasta aquí, todo es compatible con un punto de vista internalista de la explicación. Ahora bien, el principio de exclusión explicativa en su versión epistemológica nos obliga a establecer cual es la relación que existe entre esas explicaciones. Por las razones de simplicidad y unificación antes señaladas, no podemos aceptarlas a ambas; existirá, por lo tanto, una presión

para determinar la relación existente entre ambas. Si las teorías en las que se basa cada explicación no son compatibles, lo cual hace inviable una reducción conservadora, la presión derivada del principio de exclusión explicativa nos conducirá a tratar de eliminar a una de las dos.

Pero cada una de estas explicaciones en competencia nos proporciona información diferente acerca del *explanandum*: la primera de ellas, acerca de supuestas 'actitudes proposicionales' del organismo en cuestión, sobre las condiciones ambientales en las cuales esa conducta tiene lugar y sobre los significados que el organismo atribuye a la situación; la segunda explicación nos ofrece información acerca del estado del organismo en tanto que sistema físico o biológico, sus estados y procesos. Ya que, conjuntamente, nos proporcionan más información acerca de este aspecto del mundo (y, presumiblemente, acerca de otros), ¿por qué no conservar ambas explicaciones (y, consecuentemente, ambas teorías)? No parece haber una razón concluyente para pensar que el criterio de simplicidad y unificación sea *siempre* preferible al que postula las ventajas de un aumento nuestro conocimiento por medio de teorías alternativas. Para rechazar esto, un internalista que desee sostener el principio epistemológico de exclusión debería probar que las consideraciones de simplicidad y unificación son las *únicas pertinentes siempre* que tenemos explicaciones en competencia.

Nótese que no estamos afirmando que siempre es preferible un estado de menor simplicidad y unidad en nuestro conocimiento, compensado por un aumento de teorías explicativas; sólo sostenemos que, *en algunos casos*, la pérdida de simplicidad y unidad puede verse compensada por la mayor información proporcionada por teorías explicativas rivales.[163] Si, en tales casos, las consideraciones de simplicidad y unificación no resultan suficientes para fundamentar la aceptación de la versión epistemológica del principio de exclusión, hay que concluir que el único criterio que puede hacer necesaria su aceptación es la admisión previa del principio metafísico. Es importante subrayar el hecho de que, a diferencia de Sabatés, no sostenemos que las dos explicaciones alternativas acerca del mismo *explanandum* aumenten nuestra *comprensión* acerca de él. Dos razones nos llevan a rechazar esta alternativa: en primer

[163] Una situación similar se observará con respecto a las ventajas de la reducción en relación con el *desideratum* de simplicidad.

lugar, por las conocidas dificultades de lograr una caracterización mínimamente plausible del concepto de 'comprensión'; en segundo lugar porque, suponiendo que tales dificultades puedan superarse, nos parece razonable la idea de que la comprensión de los hechos del mundo está necesariamente vinculada con la determinación de las relaciones objetivas de determinación o dependencia que rigen a esos hechos y, bajo un punto de vista internalista de la explicación, las explicaciones no tienen por qué reflejar esas relaciones. Por esto último, no nos parece adecuado sostener que tales explicaciones puedan aumentar nuestra comprensión del *explanandum*. Esto implica aceptar, por lo tanto, que la unificación teórica, aunque proporciona innegables ventajas cognoscitivas, no es una condición suficiente para el logro de la comprensión.[164]

3.2. El principio de exclusión y la elección de teorías

Como hemos visto, el principio metafísico de exclusión explicativa se fundamenta en el realismo explicativo (en una concepción realista de la causalidad, si se trata de explicaciones causales). El realismo explicativo proporciona el fundamento ontológico para sostener que dos explicaciones en competencia no pueden ser ambas correctas, a menos que mantengan una relación de dependencia o sean incompletas.[165] El principio metafísico supone entonces como premisas, la aceptación de la teoría de la explicación como unificación y del realismo explicativo (aun cuando cada una de ellas puede ser suficiente para rechazar la aceptación dos explicaciones alternativas del mismo *explanandum*). La aceptación de estas dos premisas nos conduce, según Kim, a que cuando tenemos dos explicaciones en competencia nos encontramos en una situación de predicamento epistémico. Esta situación, podría decirse, es intrínsecamente inestable: existirá una presión para aumentar nuestro conocimiento del mundo, eliminando una de las explicaciones, o mostrando cómo ambas son incompletas, o señalando que una de ellas depende de la otra. La exclusión explicativa constituye, para Kim, un caso especial de la 'navaja de Occam', o principio de simplicidad: es una 'regla específica que concierne a un

[164] Bajo una concepción internalista de la explicación quizás podría bastar un *explanans* en el cual se haga referencia a correlaciones o conexiones nómicas entre conjuntos de propiedades.
[165] O, como hemos visto, que se trate de un caso de sobredeterminación.

modo importante en que se gana simplicidad en cuestiones explicativas y explica por qué esta forma de simplicidad es deseable' (1989a, p. 260).

Ahora bien, Kim no se limita a enunciar un principio general de exclusión para las explicaciones aisladas de sucesos individuales, sino que extiende el alcance de su análisis a teorías explicativas en competencia. Esta extensión no tiene una importancia epistemológica menor, dado que en la ciencia el caso más común no es encontrar explicaciones aisladas, sino con teorías completas, frecuentemente muy disímiles, cuyos dominios explicativos y predictivos se superponen. Esta extensión se advierte en el análisis de Kim relativo a los tipos de reducción de teorías, y ejemplificado por el caso de la psicología *folk* versus una neurociencia desarrollada.

El principio general que, según Kim, parece estar operando en casos como éste, es decir, de confrontación de teorías explicativas es el que sostiene que '*si una teoría se confronta con otra más explicativa, la única manera en que puede sobrevivir es siendo reducida de manera conservadora a esta última*' (1989a, p. 261. Cursivas del autor). Es decir, una de las teorías no podrá permanecer un su forma actual. Este principio, al cual podríamos denominar 'principio de exclusión explicativa para las teorías' (EET), parece constituir una extensión del principio general de exclusión explicativa 'no puede existir más de una explicación completa e independiente de un mismo suceso'.[166] Además del caso de la psicología *folk* y la neurociencia desarrollada, Kim sostiene que el principio parece operar en casos diversos, tales como la teoría de la combustión basada en el flogisto y la teoría del *ímpetus* del movimiento en su confrontación con la teoría de la oxidación y con la teoría dinámica moderna respectivamente.

Ahora bien, es pertinente preguntarse cual es el *status* epistemológico de este principio, ya que no es obvio el carácter que Kim le asigna. Si bien sugiere que el principio ha tenido una aplicación tácita en algunos casos, no es clara la importancia que le confiere en las decisiones relativas a la elección de teorías en competencia. Parecen posibles dos respuestas:

[166] Podría objetarse que en realidad no hay dos principios, sino sólo uno, ya que nunca tenemos explicaciones aisladas, sino explicaciones derivadas de teorías explicativas. No obstante, si bien la objeción parece razonable, el primer principio tendría vigencia en el caso hipotético de que nos encontráramos con un par de explicaciones aisladas acerca de un suceso, situación en la cual no necesitaríamos del segundo principio. En segundo lugar, el principio extendido es mucho más específico con respecto a la forma en que una teorías explicativa debe relacionarse con otra para no ser eliminada.

a. Es una observación general acerca del criterio de elección entre dos teorías explicativas en competencia, que presupone una cláusula *ceteris paribus* que establece la igualdad de las teorías en cuanto a todos los criterios con respecto a los cuales puede ser evaluada, criterios que son considerados independientes del poder explicativo. Esto es, un mayor poder explicativo no implica un mayor poder predictivo, ni acercamiento a la verdad, etcétera.[167]
b. Puede ser empleado como una *regla efectiva* que permite decidir el reemplazo de una teoría por otra en casos reales de elección. Esta interpretación parece estar avalada por el comentario de Kim relativo a la navaja de Occam. En este caso, el mayor poder explicativo de una teoría con respecto a otra sería un criterio *suficiente* para decidir el reemplazo de la segunda, sin necesidad de hacer referencia a ningún otro criterio de evaluación. Posiblemente, podría sostenerse que un mayor poder explicativo implica también un mayor apoyo empírico, mayor capacidad predictiva, mayor acercamiento a la verdad, etcétera.

Si el principio (EET) fuese interpretado con el alcance establecido en la afirmación a. sería difícil negar su plausibilidad (al punto que resultaría trivialmente verdadero). Si el alcance que se le atribuyera fuese el establecido por la afirmación b., la situación resultaría bastante más compleja. Podríamos preguntar, en primer lugar, si el principio (EET), bajo la interpretación expuesta en b., es suficiente como criterio para la elección de teorías.

El admitir que lo que se confronta no son explicaciones aisladas sino teorías explicativas conduce inevitablemente a situar el problema en un marco más amplio, esto es, el de cuales son los criterios que permiten decidir racionalmente el reemplazo de una teoría por otra. La pregunta pertinente sería entonces: ¿es el principio de exclusión explicativa un criterio suficientemente fuerte, por sí solo, como para proporcionar el fundamento del rechazo de una teoría científica explicativa?

Podría argumentarse aquí de la siguiente manera: sea el caso (históricamente dado) de que tenemos una explicación de un fenómeno *x* en términos de la teoría *Y*.

[167] Como el propio Kim advierte, la naturaleza de la explicación científica es un tema sobre el cual no existe en absoluto acuerdo, y que las distintas teorías de la explicación suponen relaciones distintas entre el poder explicativo de una teoría y los restantes 'méritos epistémicos' que puede acreditar. Por esta

Luego, para el mismo fenómeno *x* surge la teoría explicativa *Z*, acerca de la cual no sabemos que clase de conexión mantiene (si es que existe alguna) con la teoría *Y*. ¿Qué condiciones deberá cumplir esta coexistencia de teorías para que el principio de exclusión nos fuerce a eliminar o prescindir de una de ellas (si no puede darse el caso de que una pueda ser reducida conservadoramente a la otra)? En primer lugar, que ambas teorías sean explicativamente coextensivas, esto es, que todo lo que explica la primera teoría sea explicado por la segunda. Si hubiera alguna clase de pérdida, la ganancia epistémica por medio de la unificación de nuestro conocimiento tendría lugar a costa de una pérdida de poder explicativo, por lo cual no tendría suficiente fuerza. En segundo lugar, como Kim sostiene, que una de las teorías explique mejor que la otra los 'casos interesantes'.[168] Pero en tercer lugar, y no menos importante, podría exigirse que las teorías sean equivalentes en cuanto a otros valores epistémicos (poder heurístico, capacidad predictiva, simplicidad, compatibilidad intra y extracientífica, etcétera). Si la mejora en nuestra situación epistémica dada por un mayor poder explicativo se pierde en otros respectos, nuevamente el criterio de ganancia epistémica por medio de la unificación y simplificación carecería de suficiente fuerza como para presionar en favor de la eliminación. Dicho en otros términos, en este caso el principio de exclusión no proporciona una razón suficiente, por sí solo, como para presionar en favor de la opción por la reducción eliminativa de las teorías. Es interesante hacer notar que los casos típicamente señalados como de unificación exitosa (por ejemplo, la unificación de la mecánica celeste y terrestre por Newton), suelen ser casos en los cuales la teoría unificadora supera a sus predecesoras en *todos* los respectos relevantes, es decir, no solamente tiene mayor capacidad explicativa, sino también mayor capacidad predictiva, mayor precisión, etcétera.

Podría objetarse que estas consideraciones sólo tienen alguna fuerza si se admite que valores epistémicos tales como la simplicidad, el poder heurístico o la compatibilidad con otras teorías y otros campos, son tan importantes como el apoyo empírico (el cual incluiría el alcance explicativo, la capacidad predictiva, la

razón, no parece forzoso aceptar que el mayor poder explicativo de una teoría con respecto a otra implique automáticamente una mejora general en nuestra situación epistémica.

[168] El concepto de 'mejor desempeño' no es elucidado por Kim, aunque puede intuirse que su análisis conduciría a complicaciones adicionales.

contrastabilidad, etcétera). Pero no es necesario defender aquí la concepción según la cual en los procesos de cambio de teoría tales valores tienen igual importancia que el apoyo empírico; para los fines presentes, parece suficiente hacer notar que los mencionados procesos implican una complejidad considerable que el análisis de Kim no alcanza a reflejar, por lo que su pretensión acerca de la fuerza del principio de exclusión parece injustificada, al menos sin una considerable cualificación. Kim debería mostrar, para que el principio de exclusión alcance la fuerza que pretende, que un aumento en el poder explicativo implica un aumento generalizado en el apoyo empírico, y que éste último constituye el criterio decisivo en la elección de teorías científicas en competencia. Además, debería proporcionar en este esfuerzo una clarificación de la idea de 'mejor desempeño explicativo', crucial para la comparación de teorías. Dicho en otros términos, la carga de la prueba recae sobre Kim: debería fundamentar su pretensión de que, ante una situación de elección de teorías, o bien es racional elegir a la que tiene mayor poder explicativo, aunque la teoría rival tenga algunos méritos que la primera no tiene, o bien el mayor poder explicativo *implica* una mejora general en nuestra situación epistémica, en términos de que la teoría es más predictiva, es más aproximada a la verdad, etcétera.

Es pertinente señalar además, en relación con el análisis anterior, que el modelo de 'reducción eliminativa' al que Kim apela parece implicar consideraciones adicionales a la fuerza explicativa para que una reducción de ese tipo constituya una ganancia epistémica. En primer lugar, conviene aclarar los conceptos de 'reducción conservadora' y su alternativa, la 'reducción eliminativa'. Por 'reducción conservadora', Kim entiende una reducción que siga el modelo clásico desarrollado por Nagel (1961), en el cual una teoría es reducida a otra *vía* leyes puente que permiten que las leyes de la teoría reducida sean derivadas de las de la teoría reductora, y por medio de las cuales los conceptos de la primera sean conservados. Por 'reducción eliminativa' (o 'sustitutiva'), Kim entiende el modelo de reducción propuesto por Kemeny y Oppenheim (1956). Según Kim, en este tipo de reducción no es necesario que la teoría reducida sea lógicamente contradictoria con la teoría reductora, y ni siquiera hace falta que esté refutada; sólo hace falta que la teoría reductora explique todos los datos explicables por la teoría reducida, y que explique mejor los datos interesantes. Tampoco es necesario que existan conexiones

conceptuales o nomológicas entre las teorías, y sus vocabularios teóricos pueden ser disyuntos.

Para Kemeny y Oppenheim, para que la reducción implique un progreso en la ciencia es necesario que cumpla con ciertos requisitos, a saber

> Ciertamente se requiere que la nueva teoría cumpla el rol de la antigua, *v.g.*, que pueda explicar (o predecir) todos aquellos hechos que la antigua teoría podía manejar. En segundo lugar, no reconocemos el reemplazo de una teoría por otra como un progreso a menos que la nueva teoría se compare favorablemente con la antigua en un rasgo que podemos describir *muy aproximadamente* como simplicidad (…). Y la característica especial de la reducción es que logre todo esto y al mismo tiempo nos permita efectuar una economía en el vocabulario teórico de la ciencia (1956, p. 7. Cursivas de los autores).

En una presentación preliminar no formalizada de su modelo de reducción, Kemeny y Oppenheim proponen la siguiente caracterización de la reducción de una teoría a otra. Sean T_1 y T_2 dos teorías cuyos vocabularios teóricos son respectivamente $Voc(T_1)$ y $Voc(T_2)$. Para que se efectúe una reducción de T_2 a T_1 se requiere que

1. $Voc(T_2)$ contenga términos que no están presentes en $Voc(T_1)$.
2. Los términos contenidos en $Voc(T_2)$ que no están en $Voc(T_1)$ no sean definibles en los términos de este último.
3. T_1 pueda explicar todo lo que puede T_2.
4. T_1 no sea más compleja que T_2.

Kemeny y Oppenheim no creen que la respuesta a la pregunta '¿por qué se debería aceptar una reducción tal de una teoría a otra?' sea obvia. ¿Cuál sería la ganancia para la ciencia en que una teoría más simple sea reemplazada por una más compleja? Una respuesta posible podría ser que la teoría reductora debería ser tan simple como la teoría reducida. Pero, sostienen, esta no puede ser la respuesta completa. Si la teoría reductora es mucho más fuerte, parecería razonable permitir alguna complejidad adicional.[169] Resultaría satisfactorio, entonces, que alguna

[169] Es necesario hacer notar que la simplicidad que se obtiene con la disminución del número de términos teóricos que nos comprometemos a aceptar al adoptar la teoría reductora no implica, *per se*, una mejora

pérdida en simplicidad sea compensada por una ganancia suficiente en la fuerza explicativa y predictiva del cuerpo de teorías. Para Kemeny y Oppenheim, se necesita, por lo tanto, alguna medida que combine fortaleza (es decir, superior poder explicativo y predictivo) y simplicidad,[170] y en la cual la complejidad adicional esté balanceada por la fuerza adicional. Sugieren que este concepto combinado puede ser expresado hablando de cuán bien una teoría está *sistematizada*. Más allá de las dificultades (reconocidas por los autores) en lograr una caracterización adecuada de la noción de simplicidad, resulta evidente que la sola ganancia explicativa no parece suficiente para constituir una mejora epistémica. Kim no considera este aspecto en su análisis.

Parece haber suficientes elementos, a partir de las observaciones precedentes, para concluir que Kim no ha garantizado que el principio (EET), en su interpretación como regla efectiva, sea un criterio suficiente para decidir la elección de teorías en competencia. Resta formular ahora una observación adicional, la cual descansa en la consideración del rol *pragmático* del conocimiento científico, y que podría constituir un argumento débil en contra de una eliminación completa de la teoría reducida. La búsqueda de explicaciones de los fenómenos del mundo puede no ser considerado como un fin último, aunque a veces se sostenga que la búsqueda de conocimiento es un fin en sí mismo, en especial en relación con la ciencia básica. Si se considera, por el contrario, que la ciencia tiene también una finalidad instrumental (una formulación neutra y aceptable podría ser decir que al menos una parte de la finalidad de la ciencia es el proporcionar el conocimiento que nos permita actuar de manera eficiente para modificar el mundo), entonces las explicaciones no tienen sólo una finalidad cognoscitiva. Si esto es así, la teoría candidata a la eliminación, o que ya haya sido eliminada, puede seguir existiendo como fundamento para el diseño y desarrollo de tecnologías, aun cuando la comunidad científica ya no la considere como adecuada de acuerdo con los

---

en nuestra situación epistémica. Kim señala, en relación con el problema de los vínculos entre simplicidad y comprensión, que los sistemas más simples en términos del número de primitivos, axiomas y reglas de inferencia someten a severas pruebas a nuestra capacidad de comprensión (entendiendo en este caso por comprensión el 'hacer uso'), y señala que generalmente nos manejamos mejor con sistemas de mediana complejidad, con conceptos y reglas intuitivos, y cierta redundancia entre ellos (*cfr*. Kim, (1994)).

[170] Kemeny y Oppenheim advierten que el concepto de simplicidad necesita de 'considerable estudio', si bien existen algunos resultados parciales.

parámetros de aceptación racional vigentes. Esta situación es posible aun cuando la teoría eliminada sea considerada refutada; en este caso, su simplicidad para el diseño de tecnologías eficientes puede compensar en una medida considerable sus deficiencias como marco para la comprensión del mundo. Todo esto indica que hay una suerte de 'indeterminación' en el concepto de eliminación que Kim maneja. La eliminación de una teoría, en caso de que no pueda ser reducida a su rival, podría ser considerada una cuestión de grado, y no absoluta.

3.3. El principio de exclusión y el cambio de paradigma

Como hemos observado en el apartado 1., las consideraciones epistemológicas de Kim tienen un carácter más programático y tentativo que sus análisis respecto de los problemas de la causación mental. Este carácter esquemático de las consideraciones epistemológicas se manifiesta marcadamente en sus observaciones acerca de la concepción kuhniana de la ciencia, que examinaremos en este apartado final. Si bien este análisis constituye un aspecto de importancia marginal en el planteo de Kim, lo consideraremos a fin de completar la crítica que hemos iniciado en apartados anteriores: que la fuerza real del principio de exclusión explicativa en los casos de elección de teorías dista de ser la que le atribuye.

En una de las 'aplicaciones' de su teoría acerca de la exclusión explicativa, expuesta en su (1989a), Kim considera que ésta puede servir para explicar un aspecto 'de otro modo enigmático' de la teoría de la ciencia de Kuhn, en particular, de su concepción sobre los paradigmas. El núcleo del argumento de Kim reside en lo siguiente: los paradigmas kuhnianos son inconmensurables; no comparten metodologías, conceptos o criterios para la determinación de problemas y para la evaluación de las soluciones que se propongan. No obstante, observa Kim, a pesar de que Kuhn remarcó que los distintos paradigmas no comparten los problemas ni, en sentido estricto, los mismos datos, es plausible suponer que pueden compartir 'un dominio de cuestiones que se superponen'. De otro modo, prosigue, la teoría de los paradigmas en sí misma carecería de sentido. Y aquí es donde se plantea el problema, ya que si los paradigmas son inconmensurables en el sentido estipulado, ¿qué razones hay para no aceptarlos a todos? ¿Por qué se debería descartar el paradigma antiguo cuando se construye uno nuevo? Si es cierto que, estrictamente,

ningún paradigma es refutado, y que a pesar de ser rechazado sigue siendo útil para algunos fines explicativos y predictivos, ¿por qué no conservarlo y acumular los beneficios epistémicos que pueda brindar a los del nuevo paradigma? Si esto fuese así, continúa Kim, la teoría de la ciencia de Kuhn resultaría en una concepción acumulativa alternativa a las tradicionales teorías acumulativas de la ciencia. Pero Kuhn rechaza esta alternativa, sostiene Kim, debido a la aplicación tácita del principio de exclusión explicativa. Kuhn, finaliza Kim, considera que cada paradigma pretende brindar explicaciones completas e independientes de los problemas de su dominio, y estas explicaciones debe ser completas e independientes en relación con los restantes paradigmas en competencia.

El argumento de Kim es notoriamente inconcluyente con respecto a su pretensión de que el principio de exclusión explicativa permite esclarecer este aspecto. En particular, subestima de manera visible la importancia que el paradigma tiene para lograr la cohesión de una comunidad científica, condición *sine qua non* no sólo del progreso, sino de la existencia misma de la investigación normal. Para Kuhn, la investigación en los períodos de ciencia normal se caracteriza por ser una actividad de resolución de enigmas teóricos e instrumentales complejos determinados por un paradigma, actividad que sólo puede desarrollarse bajo el predominio monopólico del mismo paradigma. Escribe al respecto

> Los hombres cuya investigación se basa en paradigmas compartidos están sujetos a las mismas reglas y normas para la práctica científica. Este compromiso y consentimiento aparente que provoca son requisitos previos para la ciencia normal, es decir, para la génesis y la continuación de una tradición particular de la investigación científica (1962, p. 34).

A diferencia de la etapa de preciencia, en la cual los practicantes de una determinada especialidad científica se hallan inmersos en discusiones acerca de los hechos fundamentales de su campo y carecen de criterios para determinar cuales son los problemas relevantes, el predominio de un paradigma y la consiguiente etapa de ciencia normal permite a los científicos concentrarse en el tipo de investigación más esotérico de resolución de enigmas que Kuhn identifica como definitorio de la ciencia madura. Este dominio del paradigma permite a los científicos concentrarse

en determinados problemas que se supone tienen solución dentro de él; sólo su falta de ingenio, observa Kuhn, podría impedirles resolverlo.

Dado este carácter cohesivo y estructurante de la actividad científica que poseen los paradigmas en el enfoque de Kuhn, ¿qué cabría esperar si coexistieran diversos paradigmas rivales? Parece evidente que, a partir de la descripción precedente, en tales circunstancias se carecería de las condiciones básicas para el desarrollo de la ciencia normal. De alguna manera, una situación tal se asemejaría a un retorno a la etapa de preciencia: los científicos carecerían de la guía de un único paradigma, y se encontrarían probablemente envueltos en discusiones estériles (dada la inconmensurabilidad) con los proponentes de paradigmas rivales. La carencia de la orientación básica proporcionada por el paradigma se notaría especialmente en la formación de los estudiantes de la especialidad científica. Kuhn escribe con respecto a la importancia de la existencia de un paradigma dominante sobre la comunidad científica

> El estudio de los paradigmas (…) es lo que prepara principalmente al estudiante para entrar a formar parte como miembro de la comunidad científica particular con la que trabajará más tarde. Debido a que se reúne con hombres que aprenden las bases de su campo científico a partir de los mismos modelos concretos, su práctica subsiguiente raramente despertará desacuerdos sobre los fundamentos claramente expresados (1962, p. 34).

En esta descripción se advierte claramente que el carácter estructurante que Kuhn atribuye a los paradigmas es de importancia capital en la formación de quienes se inician en la práctica de una especialidad científica; la carencia de un paradigma, o la coexistencia de varios aspirantes a paradigma rivales, sólo podría tener, desde su punto de vista, efectos deletéreos sobre la organización social de los científicos y, consiguientemente, sobre la investigación normal.

Estas consideraciones ponen de manifiesto que esta interpretación alternativa de la necesidad de un único paradigma se ajusta mejor que la de Kim al espíritu de la obra de Kuhn. Nos parece, en síntesis, que la pretensión de Kim de que la exclusión de la coexistencia de paradigmas en la teoría de Kuhn se debe a una aplicación tácita del principio de exclusión es simplemente infundada.

Si el análisis precedente acerca de la justificación y los alcances epistemológicos del principio de exclusión explicativa ha sido correcto, debemos concluir que el principio general, o su extensión para el caso de las teorías explicativas, no proporciona un fundamento suficiente para rechazar (vía reducción eliminativa) teorías psicológicas en caso de que sean confrontadas con teorías neurológicas de mayor poder explicativo. Es conveniente recordar que hemos planteado esta posibilidad en forma condicional; esto es, en caso de que la exclusión causal (el problema ontológico) tenga una solución plausible.

CONCLUSIONES

En esta tesis hemos examinado el problema de la exclusión causal de los sucesos/propiedades mentales en su vinculación con la problemática de las explicaciones psicológicas. Nuestro propósito general ha sido mostrar cómo las relaciones entre los problemas de la causación mental, en particular, el de la exclusión causal, y la explicación, ya sea en su faz científica (como explicaciones provistas por la psicología científica), ya sea en su faz filosófica (como intentos de resolver el problema apelando a estrategias basadas en las doctrinas filosóficas sobre la explicación científica) son más complejas y multiformes que lo que suele reflejar la literatura. Nuestra conclusión general es que no hay una solución basada en la explicación para el problema, pero tampoco las explicaciones psicológicas se ven afectadas *in toto* por los argumentos de la exclusión. Sin embargo, convendrá revisar aquí las principales tesis que hemos sostenido y evaluar sus límites.

1. El materialismo no reduccionista, al pretender conservar la realidad, la no reducibilidad y la eficacia causal de lo mental, junto con la autonomía metodológica de la psicología, debe enfrentar el problema de la exclusión causal. No parecen existir soluciones simples que no impliquen un costo más o menos elevado en términos del abandono de ciertas intuiciones filosóficas arraigadas. Esto no necesariamente significa que el materialismo no reduccionista sea algo así como lo que en términos de Lakatos denominaríamos un 'programa de investigación degenerativo', que sólo puede tratar de superar las anomalías que se presenten, pero que ya no puede ser fructífero. Sin embargo, no sólo la cantidad sino también el tipo de respuestas destinadas a enfrentar el problema muestran la seriedad de la amenaza. Adoptar opciones como el epifenomenismo o el irrealismo retentivo equivale, en nuestra opinión, a deponer lar armas y a admitir que tal doctrina es efectivamente insostenible. Es posible que dentro de algunos años, en retrospectiva, los argumentos de la exclusión causal sean considerados como equivalentes, en relación con el materialismo no reduccionista, al argumento de la realizabilidad múltiple en relación con la teoría de la identidad: no decisivos para su refutación concluyente, pero sí suficientes para señalar deficiencias severas que lo hacen inviable en su forma actual.

2. No hemos hallado argumentos de peso como para creer que toda explicación psicológica deba verse necesariamente afectada por el problema de la exclusión. Sin embargo, el problema afecta al menos a una parte importante de las explicaciones psicológicas tanto de sentido común cuanto científicas. Este efecto se manifiesta en forma de una inconsistencia entre tales explicaciones y la conclusión de los argumentos: si éstos son correctos, las explicaciones mencionadas no pueden serlo. Los criterios de éxito explicativo que resulten apropiados para la filosofía y para la psicología no tienen necesariamente que coincidir. Podría pensarse incluso, como hemos sugerido al pasar, que ni siquiera podría hacer falta hablar, para sostener la tesis de que ha habido progreso en la psicología, de éxito explicativo. Quizás podría bastar con el éxito predictivo (para el cual puede ser suficiente la mera correlación), y el éxito tecnológico y praxiológico. En este sentido, los efectos potenciales del argumento de la exclusión causal sobre el desarrollo de los paradigmas psicológicos que sostienen la irreductibilidad y la eficacia causal de lo mental se verían notablemente atenuados. No obstante, hemos admitido que la búsqueda de explicaciones es un objetivo legítimo y fundamental de la investigación de lo mental, y que el éxito explicativo es uno de los criterios básicos por los cuales juzgamos el progreso de una ciencia. Por esta razón, hemos sostenido que los argumentos de la exclusión afectan los fundamentos filosóficos de una parte no menor de las explicaciones que la psicología actual puede proporcionar.
3. Si se acepta la ineficacia causal completa de las propiedades mentales y, pese a ello, se pretende conservar las explicaciones psicológicas, se debería ser capaz de mostrar cómo tales explicaciones, apropiadamente reformuladas, preservan la facultad de constituir el fundamento de la acción racional. Hemos planteado esta observación no como una objeción a estos modelos explicativos en sí mismos, sino como una limitación para aquellos enfoques que pretendan mantener la psicología en una forma parecida a la actual, esto es, como una disciplina autónoma (aunque no aislada), que provee explicaciones de los hechos de su dominio en sus propios términos. No carece de interés reflexionar, por otra parte, acerca del sentido que podría tener una empresa destinada al desarrollo de explicaciones que no desentrañan los verdaderos mecanismos que producen los

fenómenos y que además carecen de valor instrumental alguno. Parece plausible sospechar que tal empresa no poseería demasiada justificación epistémica o práctica.

4. Las denominadas 'estrategias deflacionistas' pueden resultar cuestionables por diversas razones; ya sea, como hemos visto, por un cuestionable otorgamiento de prioridad a la gnoseología por sobre la metafísica, o bien por el rol mismo asignado a la filosofía en el análisis del problema (en sus versiones más extremas la estrategia parece ser simplemente abandonar la reflexión filosófica sobre lo mental y dedicarse a la investigación científica). Sería una audacia injustificable sostener que muchas de las preguntas filosóficas sobre lo mental no pueden verse modificadas, y aun reemplazadas, por la investigación científica; parece igualmente extremo sostener que la reflexión filosófica no tiene demasiado que ofrecer sobre el problema. No obstante, hemos observado además que algunas de ellas (en particular, las de Baker y Burge) presentan problemas adicionales debidos a que descansan sobre supuestos sumamente discutibles y no argumentados respecto del éxito explicativo de la psicología. Estos problemas, que parecen de difícil solución, disminuyen adicionalmente la plausibilidad de la estrategia propuesta.
5. Las estrategias 'del *explanandum* dual', si bien en algunos casos parecen ser eficaces para evitar ciertos problemas (como en el caso de la propuesta de Dretske), no parecen ser capaces de preservar plenamente la causación psicofísica, excluyendo explícita o implícitamente la posibilidad de que lo mental posea efectos causales en el mundo físico. No hemos probado (ni hemos intentado hacerlo) que estas estrategias, en cualquiera de sus posibles variantes, fracasen al enfrentar el problema de la exclusión. Sin embargo, hemos considerado que cualquier propuesta de esta clase que niegue el nexo entre estados mentales y conducta física representa lo que en términos de Sabatés sería una estrategia incompatibilista, que renuncia a uno de los supuestos básicos del materialismo no reduccionista. Esto, por supuesto, puede considerarse como una deficiencia interna sólo en tanto el empleo de tal estrategia pretenda conservar la totalidad de los principios básicos de aquella doctrina. No obstante, aun en el caso de que la utilización de esta estrategia intente solamente salvar la

causación de estados mentales por otros estados mentales, parece razonable pensar que toda propuesta que renuncie a la causación de lo mental a lo físico abandona una parte fundamental de nuestra imagen del mundo, por lo que no puede ser satisfactoria.

6. La plausibilidad ontológica y el acuerdo con nuestras concepciones científicas actuales del mundo físico no pueden subordinarse a presuntas ganancias explicativas. Hemos aplicado este principio en relación con una forma reciente de dualismo, uno de cuyos argumentos es justamente el beneficio explicativo que la aceptación de tal posición implicaría. Parecería que el dualismo, en sus formas más fuertes, no puede ser considerada una opción seria en tanto tengamos buenas razones para creer que el principio de clausura causal del mundo físico es irrestrictamente verdadero. Esto equivale a decir que si, en algún momento, nuestras teorías físicas nos autorizaran a admitir que pueden existir, en ciertos casos, violaciones a este principio, el dualismo podría cobrar nueva vida. Esto implicaría admitir un mundo que es causalmente cerrado en su totalidad excepto en lo que respecta a cierta clase particular de sucesos. Si bien esta posibilidad puede parecer algo extravagante, hay analogías con el caso de otro problema básico, como es el problema del determinismo. Como algunos filósofos han observado, el determinismo no necesita ser una doctrina completamente global; puede admitirse la existencia de sucesos aleatorios o no causados en el nivel microfísico, y sin embargo sostener que todos los niveles superiores se encuentran determinados. No obstante, la situación no es equivalente al caso del dualismo: en tanto la teoría física actual nos impida aceptar violaciones a la clausura causal, esta doctrina, en sus variantes 'fuertes', no puede constituir una alternativa promisoria.
7. El principio de exclusión explicativa no puede constituir un criterio autosuficiente en favor de la eliminación de las teorías psicológicas y su reemplazo por teorías neurofisiológicas. Las ganancias explicativas deben ser contrapesadas con las pérdidas respecto de otros criterios razonables de aceptabilidad de las teorías. Aceptar esta conclusión implica admitir, como premisas, varias tesis que para algunos autores resultan inaceptables; básicamente, el propio principio de exclusión explicativa. Respecto de este

principio hemos afirmado principalmente que, en primer lugar, los argumentos más sólidos para aceptarlo son ontológicos, y no meramente epistemológicos, y, en segundo lugar que, convenientemente modificado, *puede* jugar un rol en el análisis de explicaciones y teorías explicativas en competencia.

El límite del análisis propuesto ha sido claro: ninguna de estas tesis ha supuesto ni ha intentado constituir una solución al problema de la exclusión causal. Más bien, éste ha sido considerado como un punto de partida inevitable a la hora de reflexionar sobre la causación mental y la explicación psicológica. Reiteramos que no nos parecen viables las soluciones que no impliquen un costo en términos de las intuiciones sobre lo mental a las que haya que renunciar.

Por otra parte, hay razones para no ser demasiado optimistas con respecto a la posible solución al problema de la causación mental. Una extensa lista de propuestas fallidas de solución atestigua las dificultades que se enfrentan. Si se admite que el mundo existe independientemente de nosotros y que las relaciones que regulan los fenómenos también son independientes de nosotros, entonces la relación entre lo mental y lo físico es lo que es, y el problema tiene una solución única. Sin embargo, podría ocurrir que esta solución no estuviera a nuestro alcance; simplemente, podría ocurrir que nuestro sistema cognitivo, diseñado primariamente para otras funciones a través de un largo proceso evolutivo, no fuera capaz de resolverlo. No obstante, parece claro que el desarrollo de nuestro conocimiento fáctico de lo mental no puede quedar a la espera de una solución de los problemas ontológicos básicos, y que 'soluciones' programáticas pueden constituir el fundamento para el avance de la investigación. En este intento la reflexión filosófica puede jugar un papel relevante.

Uno de los supuestos que han guiado esta investigación ha sido la convicción de que la reflexión filosófica es una actividad autónoma y diferente a la investigación científica (en este caso la investigación psicológica), pero que debe estar vinculada con ella. No existe, en este sentido, ruptura sino continuidad entre ciencia y filosofía. Las formas que, en el futuro, adopte la investigación científica acerca de lo mental seguramente serán inesperadas y sorprendentes; no puede descartarse que, en las próximas décadas, la psicología pierda su forma actual como disciplina autónoma y se produzca una integración completa con las neurociencias. En este

proceso es esperable que surjan interrogantes novedosos. Sin embargo, es muy probable que subsistan (aunque quizás en formas nuevas) las preguntas filosóficas tradicionales sobre la mente.

La subsistencia de estas preguntas es probable no sólo debido a que, como hemos observado, la respuesta a algunas de ellas parece estar más allá de la evidencia empírica. Otra causa básica parecería ser la cantidad de problemas, la mayoría de ellos fundamentales, cuya resolución parece una precondición para el éxito en esta empresa. Esto es, existe una serie de problemas tanto ontológicos cuanto gnoseológicos que suelen darse por resueltos (o, al menos, se adopta una solución programática) al enfrentar el problema de la causación mental. Entre otros pueden mencionarse: los criterios de individuación de los sucesos; la caracterización de qué es una propiedad; la naturaleza de la causalidad; las clases de determinación no causales (dependencia mereológica, superveniencia, relación determinable-determinado, dependencia de Cambridge, identidad); el concepto de lo físico y su cierre causal; el problema del significado; la caracterización de lo que es una explicación científica satisfactoria; la dependencia de la explicación de consideraciones ontológicas; el reduccionismo y el status de las 'ciencias especiales'; la relación entre la investigación científica y la reflexión filosófica; entre otros.

Muchas de estas cuestiones se han planteado en uno u otro enfoque del problema. Prácticamente ninguno de los problemas planteados parece más sencillo de resolver que el propio problema de la exclusión. ¿Sería plausible exigir a una propuesta de solución que adopte una posición sólida, consistente y fundamentada a todas estas cuestiones? Parecería no sólo irrazonable, sino también impracticable.[171] Y si esta exigencia es irrealizable, parecería razonable concluir que cualquier propuesta de solución al problema no podrá ser más fuerte que cualquiera de los eslabones constituidos por las respuestas programáticas de solución a los problemas mencionados.

La búsqueda de una explicación apropiada para la relación entre lo mental y lo físico ha perdido su carácter de misterio: disponemos tanto de herramientas conceptuales cuanto empíricas para enfrentarla. Sin embargo, parece plausible la

sospecha de que este problema nos seguirá acompañando, en una u otra forma, durante mucho tiempo.

[171] Pese a que exigencias de esta clase son planteadas a veces por algunos filósofos. Bunge, por ejemplo, exige que las 'filosofías de la mente' se enmarquen dentro de una ontología global (*cfr*. al respecto Bunge y Ardila (1988)).

BIBLIOGRAFÍA

Armstrong, David M. (1978), *Los universales y el realismo científico*, México, UNAM, 1988.

Ajzen, Icek y Martin Fishbein (1980), *Understanding Attitudes and Predicting Social Behavior*, Englewood Cliffs, Prentice Hall.

Baker, Lynn (1993), 'Metaphysics and Mental Causation', en J. Heil y A. Mele (eds.), 1993.

Bechtel, William (1988) *Filosofía de la Mente. Una panorámica para la ciencia cognitiva*, Madrid, Tecnos, 1991.

Beckermann, Ansgar (1992), 'Reductive and Nonreductive Physicalism', en A. Beckermann, H. Flohr, & J. Kim (eds) *Emergence or Reduction?: Prospects for Nonreductive Physicalism*, Berlín, De Gruyter.

Berger, Peter y Thomas Luckmann (1966), *La construcción social de la realidad*, Buenos Aires, Amorrortu, 1983.

Bieri, Peter (1992), 'Trying Out Epiphenomenalism', *Erkenntnis*, 36, 3, pp. 283-309.

Block, Ned (1990), 'Can the Mind Change the World?', en C. y G. Macdonald (eds.), 1995.

Broncano, Fernando (comp.) (1995), *La mente humana*, Madrid, Trotta.

Broncano, Fernando (2000), *Mundos artificiales. Filosofía del cambio tecnológico*, México, Paidós-UNAM.

Bunge, Mario (1981), *El problema mente-cerebro. Un enfoque psicobiológico*, Madrid, Tecnos, 1985.

Bunge, Mario (1983), *La investigación científica. Su estrategia y su filosofía*, Barcelona, Ariel, 1989.

Bunge, Mario y Rubén Ardila (1988), *Filosofía de la Psicología*, Barcelona, Ariel.

Bunzl, Martin (1979), 'Causal Overdetermination', *Journal of Philosophy*, 76, pp. 134-150.

Burge, Tyler (1986), 'El individualismo y la psicología', en E. Rabossi (comp.) (1995), *Filosofía de la mente y ciencia cognitiva*, Barcelona, Paidós.

Burge, Tyler (1993), 'Mind-Body Causation and Explanatory Practice', en J. Heil y A. Mele (eds.), 1993.

Churchland, Paul (1988), *Materia y conciencia. Introducción contemporánea a la Filosofía de la Mente*, Barcelona, Gedisa, 1992.

Comesaña, Manuel (1995), *Razón, verdad y experiencia. Un análisis de sus vínculos en la epistemología contemporánea, con especial referencia a Popper*, Mar del Plata, Universidad Nacional de Mar del Plata.

Crane, Tim (1995), 'The Mental Causation Debate (Mental Causation I)', *Aristotelian Society Supplement*, 69, pp. 211-236.

Cummins, Robert (1983), *The Nature of Psychological Explanation*, Cambridge and London, MIT Press/Bradford Books.

Davidson, Donald (1963), 'Acciones, razones y causas', *Journal of Philosophy*, 60. Reimpreso en Davidson (1980).

Davidson, Donald (1970), 'Sucesos mentales', en L. Foster y J. W. Swanson (eds.), *Experience and Theory*, Mass., University of Massachusetts Press y Duckworth, 1970. Reimpreso en Davidson (1980).

Davidson, Donald (1980), *Ensayos sobre acciones y sucesos,* Barcelona, UNAM-Crítica, 1995.

Dennet, Daniel (1991), *La conciencia explicada*, Barcelona, Paidós, 1995.

Dretske, Fred (1990), 'Does Meaning Matter?', en C. y G. Macdonald (eds.), 1995.

Dretske, Fred (1995), 'Reply: Causal Relevance and Explanatory Exclusion', en C. y G. Macdonald (eds.), 1995.

Engel, Pascal (1988), '¿Puede la psicología cognitiva apelar a la psicología ordinaria?', en Pascal Engel (comp.) (1988), *Psicología ordinaria y ciencias cognitivas*, Barcelona, Gedisa, 1993.

Ezquerro, Jesús (1995), 'Teorías de la arquitectura de lo mental', en F. Broncano (comp.) (1995).

Fernández Acevedo, Gustavo (1998), 'Realismo e internalismo en el modelo nomológico-deductivo', en AAVV, *Estudios sobre epistemología y ciencias sociales*, Mar del Plata, Editorial Martín - Universidad Nacional de Mar del Plata.

Friedman, Michael (1974), 'Explanation and Scientific Understanding', *Journal of Philosophy*, 71, 1, pp. 5-19. Reimpreso en Joseph Pitt (ed.) (1988), *Theories of Explanation*, Nueva York, Oxford University Press.

Fodor, Jerry (1968), *La explicación psicológica. Introducción a la filosofía de la psicología*, Madrid, Cátedra, 1980.

Fodor, Jerry (1989) 'Making Mind Matter More', *Philosophical Topics*, 17, pp. 59-80.

Gardner, Howard (1985), *La nueva ciencia de la mente. Historia de la revolución cognitiva*, Barcelona, Paidós, 1987.

Glymour, Clark (1999), 'A Mind Is a Terrible Thing to Waste - Critical Notice: Jaegwon Kim, "Mind in a Physical World"', *Philosophy of Science*, 66, 3, pp. 455-471.

Goldman, Alvin (1969), 'The Compatibility of Mechanism and Purpose', *Philosophical Review*, 78, pp. 468-482.

Green, Christopher (1997), 'Mental Causation and Connectionism', *Convention of the Canadian Psychological Association*.

Green, Celia y Grant Gillett (1995), 'Are Mental Events Preceded by Their Physical Causes?', *Philosophical Psychology*, 8, 4, pp. 333-340.

Guttenplan, Samuel (ed.) (1994), *A Companion to the Philosophy of Mind*, Oxford, Blackwell.

Haack, Susan (1995), *Evidence and Inquiry. Toward Reconstruction in Epistemology*, Oxford, Blackwell.

Hardcastle, Valerie Gray (1998), 'On the Matter of Minds and Mental Causation', *Philosophy and Phenomenological Research*, 58, pp. 1-25.

Harré, Rom, David Clarke y Nicola De Carlo (1985), *Motivos y mecanismos. Una introducción a la psicología de la acción*, Barcelona, Paidós, 1989.

Hart, William D. (1994), 'Dualism', en S. Guttenplan (ed.) (1994).

Heil, John (1991), 'On the Cutting Edge: Philosophical Perspectives on Mental Causation', *Philosophical Papers*, 20, 2, pp. 113-137.

Heil, John (1992), 'Mentality and Causality', *Topoi*, 11, pp. 103-110.

Heil, John, y Alfred Mele (eds.) (1993), *Mental Causation*, New York, Oxford University Press.

Hempel, Carl (1942), 'El papel de las leyes generales en historia', en C. Hempel (1965b).

Hempel, Carl (1958), 'La lógica de la explicación científica', en C. Hempel (1965b).

Hempel, Carl (1965a), 'Aspectos de la explicación científica', en C. Hempel (1965b).

Hempel, Carl (1965b), *La explicación científica. Estudios sobre la filosofía de la ciencia*, Barcelona, Paidós, 1988.

Hempel, Carl (1966) *Filosofía de la ciencia natural*, Madrid, Alianza, 1979.

Hewstone, Miles (1989), *La atribución causal*, Barcelona, Paidós, 1992.

Honderich, Ted (1984), 'Is the Mind Ahead of the Brain? Benjamin Libet's Evidence Examined', en http://www.ucl.ac.uk/~uctytho/libet1.htm. Publicado anteriormente bajo el título 'The Time of a Conscious Sensory Experience and Mind-Brain Theories', en *The Journal of Theoretical Biology*.

Horgan, Terence (1997), 'Kim on Mental Causation and Causal Exclusion', *Philosophical Perspectives*, 11, pp. 165-184.

Horgan, Terence y John Tienson (1993), 'Levels of Description in Nonclassical Cognitive Science', *Philosophy Supplement*, 34, pp. 159-188.

Horowitz, Amir (1999), 'Is There a Problem in Physicalist Epiphenomenalism?', *Philosophy and Phenomenological Research*, LIX, 2, pp. 421-434.

Jackson, Frank (1996), 'Mental Causation', *Mind*, 105, pp. 377-413.

Jackson, Frank y Philip Pettit (1990), 'Program Explanation: a General Perspective', *Analysis*, 50, pp. 107-117.

Kemeny, John y Paul Oppenheim (1956), 'On Reduction', *Philosophical Studies*, 7, pp. 6-19.

Kim, Jaegwon (1974), 'Noncausal Connections', *Noûs*, 8, pp. 41-52. Reimpreso en J. Kim (1993b).

Kim, Jaegwon (1976) 'Events as Property Exemplifications', en Miles Brand y Douglas Walton (eds.), *Action Theory*, Dordrecht, Reidel. Reimpreso en J. Kim (1993b).

Kim, Jaegwon (1979), 'Causality, Identity, and Supervenience in the Mind-Body Problem', *Midwest Studies in Philosophy*, 4, pp. 31-49.

Kim, Jaegwon (1984a), 'Supervenience and Supervenient Causation', *Southern Journal of Philosophy*, 22, pp. 45-56.

Kim, Jaegwon (1984b) 'Epiphenomenal and Supervenient Causation', *Midwest Studies in Philosophy*, 9, pp. 257-270. Reimpreso en J. Kim (1993b).

Kim, Jaegwon (1987) 'Explanatory Realism, Causal Realism, and Explanatory Exclusion', en P. French, T. Uehling and H. Wettstein (eds.) *Midwest Studies in Philosophy,* vol. 12, Notre Dame, University of Notre Dame Press.

Kim, Jaegwon (1989a) 'Mechanism, Purpose, and Explanatory Exclusion', *Philosophical Perspectives*, 3, *Philosophy of Mind and Action Theory*, James E. Tomberlin (ed.), Atascadero, Cal., Ridgeview Publishing Company. Reimpreso en J. Kim (1993b).

Kim, Jaegwon (1989b), 'The Myth of Nonreductive Materialism', *Proceedings and Addresses of the American Philosophical Association*, 63, pp. 31-47. Reimpreso en J. Kim (1993b)

Kim, Jaegwon (1990), 'Explanatory Exclusion and the Problem of Mental Causation', en Cynthia y Graham Macdonald (eds.), 1995.

Kim, Jaegwon (1991) 'Dretske on How Reasons Explains Behavior', en Brian McLaughlin (ed.), *Dretske and his Critics*, Oxford, Basil Blackwell, 1991. Reimpreso en J. Kim (1993b).

Kim, Jaegwon (1993a), 'The Nonreductivist's Troubles with Mental Causation', en John Heil y Alfred Mele (eds.), *Mental Causation*, Oxford, Oxford University Press, 1989. Reimpreso en J. Kim (1993b).

Kim, Jaegwon (1993b), *Supervenience and Mind*, Cambridge, Cambridge University Press.

Kim, Jaegwon (1994) 'Explanatory Knowledge and Metaphysical Dependence', en E. Villanueva (ed.), *Philosophical Issues*, 5, Truth and Rationality, Atascadero, Ca., Ridgeview Publ. Co., 1994.

Kim, Jaegwon (1995), 'Mental Causation in Searle's "Biological Naturalism"', *Philosophy and Phenomenological Research*, 55, 1, pp. 189-194.

Kim, Jaegwon (1998), *Mind in a Physical World. An Essay on the Mind-Body Problem and Mental Causation*, Cambridge, Mass., The MIT Press.

Kuhn, Thomas (1962), *La estructura de las revoluciones científicas*, Buenos Aires, Fondo de Cultura Económica, 1990.

Laudan, Larry (1977), *El progreso y sus problemas. Hacia una teoría del crecimiento científico*, Madrid, Encuentro, 1986.

Loeb, Louis (1974), 'Causal Theories and Causal Overdetermination', *Journal of Philosophy*, 71, pp. 525-544.

Loewer, Barry (1995), 'Mind-body problem', en J. Kim y E. Sosa (eds.), *A Companion to Metaphysics*, Oxford, Blackwell, 1995.

Lowe, E. Jonathan (1999), 'Self, Agency and Mental Causation', *Journal of Consciousness Studies*, 6, 8-9, pp. 225-239.

Macdonald, Cynthia y Graham Macdonald (1986), 'Mental Causes and Explanation of Action', *Philosophical Quarterly*, 36, pp. 145-158.

Macdonald, Cynthia y Graham Macdonald (eds.) (1995), *Philosophy of Psychology. Debates on Psychological Explanation, Volume One*, London, Blackwell.

Malcolm, Norman (1968), 'The Conceivability of Mechanism', *Philosophical Review*, 77, pp. 45-72. Reimpreso en *Free Will*, Gary Watson (ed.), Oxford University Press, 1982.

Marr, David (1977), 'La inteligencia artificial: un punto de vista personal', en Margaret Boden (comp.) (1990), *Filosofía de la inteligencia artificial*, México, Fondo de Cultura Económica, 1994.

Marr, David (1982), *La visión*, Madrid, Alianza, 1985.

Marras, Ausonio (1993), 'Psychophysical Supervenience and Nonreductive Materialism', *Synthese*, 95, pp. 275-304.

Marras, Ausonio (1994), 'Nonreductive Materialism and Mental Causation', *Canadian Journal of Philosophy*, 24, 3, pp. 465-493.

Marras, Ausonio (1997a), 'The Causal Relevance of Mental Properties', *Philosophia*, 25, 1-4, pp. 389-400.

Marras, Ausonio (1997b), 'Metaphysical Foundations of Action Explanation', en G. Holmström Hintikka y R. Tuomela (eds.), *Contemporary Action Theory*, vol. 1, Dordrecht, Kluwer.

Marras, Ausonio (1998), 'Kim's principle of Explanatory Exclusion', *Australasian Journal of Philosophy*, 76, pp. 439-451.

Marras, Ausonio (2000), 'Critical Notice of *Mind in a Physical World* by Jaegwon Kim', *Canadian Journal of Philosophy*, 30, pp. 137-160.

McClamrock, Ron (1991), 'Marr's Three Levels: A Re-evaluation', en http://www.albany.edu/~ron/papers/marrlevl.html (publicado en *Minds and Machines*, Mayo de 1991).

McLaughlin, Brian (1989), 'Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical', *Philosophical Perspectives*, 3, pp. 109-135.

McLaughlin, Brian (1994), 'Epiphenomenalism', en Guttenplan, S. (ed.), 1994.

Montgomery, Richard (1995), 'Non-Cartesian Explanations Meet the Problem of Mental Causation', *Southern Journal of Philosophy*, 33, 2, pp. 221-241.

Nagel, Ernest (1961), *La estructura de la ciencia*, Barcelona, Paidós, 1991.

Newton-Smith, William H. (1981), *La racionalidad de la ciencia*, Barcelona, Paidós, 1987.

Pérez, Diana (1996), 'Variedades de superveniencia', *Manuscrito,* XIX, 2, pp. 165-99.

Pérez, Diana (2000), *La mente como eslabón causal*, Buenos Aires, Catálogos.

Piaget, Jean (1963), 'La explicación en psicología y el paralelismo psicofísico', en P. Fraisse, J. Piaget y M. Reuchlin (comps.), *Historia y método de la psicología experimental*, Buenos Aires, Paidós, 1972.

Pinker, Steven (1997), *Cómo funciona la mente*, Buenos Aires, Destino, 2001.

Popper, Karl (1934), *La lógica de la investigación científica*, Buenos Aires, REI, 1985.

Popper, Karl (1972), 'El objeto de la ciencia', en K. Popper (1972), *Conocimiento Objetivo*, Madrid, Tecnos, 1992.

Popper, Karl, y John Eccles (1977), *El yo y su cerebro*, Barcelona, Labor, 1993.

Putnam, Hilary (1967), 'Psychological Predicates', en W. H. Capitan y D. D. Merril (eds.), *Art, Mind and Religion*, Pittsburgh, University of Pittsburgh Press (reimpreso como 'The Nature of Mental States'). ('La naturaleza de los estados mentales', México, Cuadernos de Crítica, 1981).

Quintanilla, Miguel Angel (1988), *Tecnología: un enfoque filosófico*, Buenos Aires, Fundesco-Eudeba.

Rabossi, Eduardo (1995a), 'La tesis de la identidad mente-cuerpo', en F. Broncano (comp.) (1995).

Rabossi, Eduardo (1995b), 'Notas sobre el no reduccionismo y la realizabilidad variable', *Análisis Filosófico*, XV, 1 y 2, pp. 167-179.

Ruben, David (1990), *Explaining Explanation*, Londres, Routledge.

Sabatés, Marcelo (1996), 'Kim on the Metaphysics of Explanation', *Manuscrito*, XIX, 2, pp. 93-110.

Sabatés, Marcelo (1997), 'Should a Cognitive Psychologist Worry About the Causal Inefficacy of the Mental?', en B. Niggemeyer (ed.), *The Cognitive Level*, Duisburg, LAUD Verlag.

Sabatés, Marcelo (2001), 'Varieties of Exclusion', *Theoria*, Vol. 16, Nº 40, pp. 13-42.

Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton, Princeton University Press.

Searle, John (1992), *El redescubrimiento de la mente*, Barcelona, Crítica, 1996.

Skidelsky, Liza (1996) 'La cuestión de los niveles explicativos de una teoría cognitiva', en M. Velasco y A. Saal (comps.), *Epistemología e Historia de la Ciencia*, Córdoba, Universidad Nacional de Córdoba, pp. 290-296.

Thomasson, Amie (1998), 'A Nonreductivist Solution to Mental Causation', *Philosophical Studies*, 89, pp. 181-95.

Van Fraassen, Bas (1980), *La imagen científica*, México, UNAM, 1996.

Van Gulick, Robert (1993), 'Who's In Charge Here? And Who's Doing All the Work?', en J. Heil y A. Mele (eds.), 1993.

Vicente, Agustín (2002), 'The Dual *Explanandum* Strategy', *Crítica*, 34, 101, pp. 73-96.

Von Wrigth, Georg H. (1971), *Explicación y comprensión,* Madrid, Alianza, 1987.