

MECANISMO, PROPOSITO Y EXCLUSION EXPLICATIVA*

JAEGWON KIM

Quiero reabrir la discusión acerca de si una misma porción de comportamiento, digamos una acción tal como subir una escalera, puede recibir tanto una explicación "mecanicista" en términos de procesos y leyes fisiológicos como una explicación "por propósitos" [*"purposive" explanation*], en términos de "razones" (por ejemplo, metas y creencias). En un artículo publicado en 1968,¹ Norman Malcolm defendió una respuesta negativa. Argumentó que una vez que una acción ha sido explicada estableciendo sus antecedentes causales fisiológicos, se nos cierra la posibilidad de explicarla citando las razones del agente, es decir, sus creencias, deseos, intenciones y cosas semejantes. Alvin Goldman contestó inmediatamente a Malcolm,² argumentando que las explicaciones mecanicistas y "por propósitos" son compatibles, y que, de hecho, podemos caracterizar un tipo de situación en la que uno y el mismo comportamiento puede considerarse explicable tanto fisiológica como racionalmente.

Quiero reabrir este debate no sólo porque hay más cosas que pueden decirse sobre el tema sino también, principalmente, porque el tema tiene importantes implicaciones sobre algunos problemas que en la actualidad revisten mucho interés en la filosofía de la mente. Una apreciación adecuada de los problemas y opciones metodológicas más amplias involucradas nos ayudará, creo, a aclarar algunas de las cuestiones de la controversia actual. Tal como veremos, la pregunta acerca de la compatibilidad explicativa nos lleva a preguntas más generales sobre la posibilidad de *explicaciones múltiples de un único explanandum*, y la relación entre dos teorías

* Traducción: Cecilia Hidalgo y Juan Rodríguez Larreta

¹ "The Conceivability of Mechanism", *Philosophical Review* 77 (1968): 45-72. Reimpreso en *Free Will* de Gary Watson (comp.) (Oxford: Oxford University Press, 1982). Las referencias de página a este artículo corresponden a la reimpresión en la compilación de Watson. Para la defensa anterior de una posición similar a la de Malcolm véase "Determinism", de Alasdair C. MacIntyre, *Mind* 66 (1957): 28-41.

² "The Compatibility of Mechanism and Purpose", *Philosophical Review* 78 (1969): 468-82; también en su *A Theory of Human Action* (Englewood Cliffs, N. J.: Prentice-Hall, 1970), pp. 157-65. Para otra defensa de la posición compatibilista y una crítica de MacIntyre, véase Daniel C. Dennett, "Mechanism and Responsibility", en *Brainstorms* de Dennett (Montpelier, Vt.: Bradford Books, 1981). G. H. von Wright defiende una "solución con dos *explananda*" en *Explanation and Understanding* (Ithaca: Cornell University Press, 1971), pp. 118-31.

explicativas distintas que cubren dominios de fenómenos superpuestos. En última instancia, se verá que estos temas surgen de algunas suposiciones básicas acerca de la epistemología y de la metafísica de la explicación y, en particular, de la cuestión del “realismo” acerca de las explicaciones.

Lo que Malcolm denomina “explicación por propósitos” de una acción es algo que se ajusta al patrón familiar “creencia-deseo” de explicación de la acción. Tal explicación explica una acción especificando la “razón por la cual” el agente hizo lo que hizo, es decir, indicando lo que quería realizar y lo que consideró una manera óptima de realizar lo que quería. Nos referiremos a tales explicaciones como “explicaciones racionalizadoras” [*rationalizing explanations*]; estas explicaciones nos proporcionan las consideraciones del agente, explícitas o implícitas, que racionalizan las acciones a ser explicadas. No necesitamos pensar a las explicaciones creencia-deseo como el único tipo de explicaciones racionalizadoras (“Le pegué porque insultó a mi esposa”); ni precisamos pensar a las racionalizaciones como el único modo de explicación de la acción en la psicología ordinaria [*vernacular psychology*]. Sin embargo, las explicaciones creencia-deseo parecen ocupar un lugar central en nuestra comprensión cotidiana de lo que nosotros y nuestros semejantes hacemos y se podría pretender correctamente, pienso, que constituyen el modo básico de comprender las acciones en la psicología intencional. Más aun, la pregunta acerca de si tales explicaciones pueden o no coexistir con las explicaciones fisiológicas del comportamiento se generalizará, sin duda, a otros modos de explicación intencional, y la manera como respondamos esta pregunta tendrá implicaciones directas para el debate actual acerca de la relación entre la psicología ordinaria “popular” (“*folk*”) y la ciencia sistemática del comportamiento humano, sea este último considerado una “ciencia cognitiva” de relativamente alto nivel o una “neurociencia” de bajo nivel.³

Pero este problema de la relación entre la psicología ordinaria y la sistemática puede en sí mismo ser generalizado: ¿cuál es la relación entre dos teorías explicativas (en especial, dos teorías sucesivas) de los mismos fenómenos? ¿Pueden los mismos fenómenos ser explicados correctamente por dos teorías diferentes? ¿Podemos aceptar dos teorías diferentes, cuando ambas pretenden proporcionar explicaciones independientes de los mismos datos? Supongamos que aceptamos ya sea la tesis de la com-

³ Véase, por ej., Paul Churchland, *Scientific Realism and the Plasticity of Mind* (Cambridge: Cambridge University Press, 1979); Stephen P. Stich, *From Folk Psychology to Cognitive Science* (Cambridge: The MIT Press, 1983), Terence Horgan y James Woodward, “Folk Psychology Is Here To Stay”, *Philosophical Review* 94 (1985), 197-226.

patibilidad, o la tesis de la incompatibilidad, en lo que concierne a ambos tipos de explicación de la acción. ¿Podría generalizarse alguna de las dos posiciones? ¿Es cierto *en general* que puede darse más de una explicación —o más de un tipo de explicación— de un acontecimiento [*event*]* dado? ¿O es cierto en general que de ningún acontecimiento (quizá, de nada) puede ofrecerse más de una explicación? ¿Hay condiciones generales bajo las cuales las explicaciones se *excluyen* entre sí?

Ahora pondré mis cartas sobre la mesa. Acerca de la cuestión de la compatibilidad de las explicaciones de la acción, pienso que Malcolm está fundamentalmente en lo correcto, aunque, como veremos, esto no necesariamente muestra que Goldman esté equivocado. Argumentaré que, aun cuando las críticas que se han planteado contra Malcolm señalan algunas posibilidades interesantes y es preciso tomarlas en cuenta, ellas no refutan lo que considero el corazón de los argumentos de Malcolm, y que con las suposiciones auxiliares especiales que Malcolm parece aceptar, las explicaciones racionalizadoras y fisiológicas se excluyen entre sí. Mis consideraciones centrales no dependerán de ningún rasgo especial de las razones y las causas, o de la mente y la materia, sino que involucran en cambio ciertas reflexiones generales acerca de la naturaleza de la explicación y de la causación —en particular, los tipos de situación en los que las explicaciones con *explanantia* mutuamente consistentes pueden no obstante competir entre sí. Esto me llevará a formular lo que denominaré “el principio de exclusión explicativa”, algo que temo muchos considerarán absurdamente fuerte e inaceptable. En términos aproximados, este principio dice lo siguiente: a ningún acontecimiento puede dársele más que una explicación *completa e independiente*. Lo que “completa” e “independiente” pueden significar en este contexto es obviamente importante, y mi discusión tomará en cuenta la necesidad de aclarar estas nociones. Debería decir ahora mismo, sin embargo, que no ofreceré definiciones generales de estas nociones sino que me apoyaré más bien en la discusión de casos específicos a fin de generar sentidos razonablemente compatibles [*cohesive*] para estos términos. Antes bien, mi estrategia será la siguiente: discutiré mi posición principalmente con respecto a las *explicaciones causales*, proponiendo al mismo tiempo algunas consideraciones generales que, espero, harán que el principio de exclusión parezca una restricción plausible sobre las explicaciones en general.

* Usaremos indistintamente “acontecimiento” y “evento” como traducción de “*event*”. (N. del T.)

1. El argumento de Malcolm

Aun cuando tanto la forma exacta de la explicación fisiológica del comportamiento, como la cuestión de cómo vemos exactamente la estructura de una explicación racionalizadora no son en última instancia importantes (como lo serían si nuestros resultados fueran a ser de interés general), será útil tener algunos puntos de referencia fijos. Para ello podemos simplemente detenernos en la concepción de Malcolm sobre el tema. El considera que una explicación fisiológica del comportamiento tiene la forma hempeliana familiar del esquema de “cobertura legal” o “nomológico-deductivo” (“N-D”):

(N) Siempre que un organismo de estructura *S* está en un estado neurofisiológico *g*, emitirá un movimiento *m*.

El organismo *O* de estructura *S* estaba en el estado neurofisiológico *g*.

Por tanto, *O* emitió *m*.

Para Malcolm una explicación racionalizadora tiene la siguiente forma (también familiar):

(R) Siempre que un organismo *O* tiene la meta *M* y cree que se requiere el comportamiento *C* para producir *M*, *O* emitirá *C*.

O tenía *M* y creía que se requería *C* para *M*.

Por tanto, *O* emitió *C*.

La cuestión de si (R), también, es un esquema N-D (Malcolm considera que no lo es) no desempeñará ningún papel en la discusión que sigue, aun cuando precisamos suponer, con Malcolm, que (R) representa el complejo meta-creencia como una “causa” o “condición” del acaecer del comportamiento. Esto es importante: tal como se tornará más claro luego, la incompatibilidad entre estas explicaciones resalta de la manera más completa cuando ambas son interpretadas como explicaciones causales, como intentos de proporcionar condiciones causales de las que surgieron la acción o el comportamiento.⁴

Antes de presentar su argumento de incompatibilidad, Malcolm trata de excluir la posibilidad de que una explicación racionalizadora de la for-

⁴ Esto significa que una manera en la que se podría tratar de eliminar la incompatibilidad es interpretar la explicación racionalizadora como un modo fundamentalmente *no causal* de comprender las acciones. Creo que en última instancia éste es el enfoque que deberíamos adoptar: una explicación racionalizadora tendría que ser vista como una *evaluación normativa* de una acción en el contexto de los estados intencionales del agente que son relevantes. Para una elaboración de esta idea véase mi “Self-Understanding and Rationalizing Explanations”, *Philosophia Naturalis* 21 (1984): 309-20.

ma (R) sea “menos básica que” y “dependiente de”⁵ las explicaciones fisiológicas. Pienso que el intento de Malcolm no tiene éxito, por diversas razones.⁶ Pero que Malcolm tenga o no éxito aquí es menos interesante para nuestros propósitos que el porqué del intento. Recordemos que la formulación provisional del principio de exclusión explicativa que ofrecí antes sólo prohíbe que haya más de una explicación completa e independiente de un fenómeno dado. Aun cuando Malcolm no dice por qué toma la cuestión de la dependencia explicativa en este punto, no hay que ir muy lejos para encontrar una razón [*rationale*]: si la explicación racionalizadora depende de la explicación fisiológica en un sentido apropiado (por ejemplo, por ser reducible a ella), entonces en verdad aquí hay sólo una explicación, y no surge la cuestión de la compatibilidad explicativa. Ambas explicaciones podrían coexistir pacíficamente, pero la paz se lograría a un precio: ya no hay explicaciones independientes. Se habría mostrado que la eficacia explicativa de una deriva de la otra y en última instancia la explicación fisiológica tendría que ser considerada como narrando una historia más profunda y abarcativa acerca de cómo surgió el comportamiento.

¿Cuál es entonces, tal como lo expresa Malcolm, “la relación lógica exacta entre las explicaciones neurológicas y por propósitos del comportamiento”?⁷ El pregunta: “¿Pueden las explicaciones de ambos tipos ser verdaderas acerca de la misma porción de comportamiento en una y la misma ocasión?”.⁸ Este es el problema de la compatibilidad explicativa. Pero habrán advertido que ambos esquemas explicativos, (N) y (R), no tienen, estrictamente hablando, el mismo *explanandum*; sus conclusiones explicativas son diferentes. Una habla de la “emisión de (el movimiento corporal) *m*” y la otra de la “emisión de (el comportamiento) *C*”. Dada esta aparente diferencia entre los *explananda*, parecería que no puede surgir la cuestión de la compatibilidad. Malcolm se percata de esta “solución del *explanandum* dual” (como podríamos denominarlo) y responde de la siguiente manera:⁹

⁵ Malcolm, “The Conceivability of Mechanism”, p. 131.

⁶ En primer lugar, la noción de “dependencia” usada es demasiado estrecha y en el mejor de los casos parece caracterizar un subcaso especial; en segundo lugar, su argumentación hace uso de suposiciones especiales que necesitan justificación y de lo que parecen ser rasgos locales del caso particular en manos. Para una discusión del argumento de Malcolm véase “Neurophysiological Laws and Purposive Principles” de William L. Rowe, *Philosophical Review* 81 (1971): 502-508.

⁷ Malcolm, “The Conceivability of Mechanism”, p. 132.

⁸ *Ibid.*

⁹ *Ibid.*, p. 133.

Tomemos el ejemplo del hombre subiendo la escalera para recoger su sombrero del techo. Esta explicación relaciona su subida con su intención. Una explicación neurofisiológica de su subida no diría nada acerca de su intención pero conectaría sus movimientos en la escalera con cambios químicos en los tejidos corporales o con las descargas neuronales. ¿Interfieren entre sí ambas explicaciones?

Creo que habría una colisión entre ambas explicaciones si se las ofreciera como explicaciones de uno y el mismo acaecimiento de un hombre subiendo una escalera.

Aún no queda claro exactamente cómo responde esto el problema del *explanandum* dual. Malcolm parece pensar que hay aquí un acontecimiento concreto —el movimiento ascendente de la escalera por parte del hombre, representado por ambas conclusiones— que hace las veces de *explanandum* compartido de ambas explicaciones. En nuestra discusión supondremos que ésta es la situación (Goldman no cuestiona esta suposición); si no lo hiciéramos, el problema de la incompatibilidad explicativa podría replantearse, aunque esto acarrearía complicaciones.¹⁰ Podemos simplemente señalar aquí que aun cuando los dos enunciados *explanandum* no son equivalentes o sinónimos, hay un sentido evidente en el que “describen” uno y el mismo acontecimiento, el mismo acaecimiento concreto, y podríamos considerar que el problema de la compatibilidad se plantea con respecto a este acontecimiento cualquiera sea la manera como se lo describa. Por ahora, en tanto exista, no importará cuál es exactamente este *explanandum* compartido.

El argumento de Malcolm en favor de la pretensión de que no pueden aceptarse ambas explicaciones parece hacer uso de la siguiente suposición:¹¹

(I) Si el acontecimiento *A* es nomológicamente suficiente para el acaecimiento del acontecimiento *E*, entonces ningún acontecimiento completamente distinto de *A* es necesario para que acaezca *E*.

Al usar este principio, Malcolm parece argüir de la siguiente manera: supongamos que hay una explicación fisiológica del subir una escalera por parte de un hombre que se ajusta al esquema (N) presentado anteriormente. Esta explicación muestra a cierto acontecimiento fisiológico (“el estado neurofisiológico *g*”) como siendo nomológicamente suficiente

¹⁰ Creo que hay mucho para decir en favor del enfoque “Dos *explananda*” en el caso de la explicación del comportamiento (véase mi “Self-Understanding and Rationalizing Explanations”). Acerca de este tema soy bastante sumario aquí pues mi interés principal concierne a la cuestión general de la compatibilidad explicativa.

¹¹ Tal como lo formula Goldman (con pequeños cambios de redacción).

para el comportamiento. Si este acontecimiento fisiológico es en verdad *suficiente* para la subida, la subida debería acaecer aun cuando no acaeciera ningún otro acontecimiento (tal como creencias y deseos). Es decir, no sería necesario ningún *otro* acontecimiento para que se dé el acaecimiento de la subida, y la explicación fisiológica en sí misma será considerada completa y suficiente como explicación del comportamiento. Una vez que sabemos que está presente la condición fisiológica, podemos confiar plenamente en que acaecerá la subida de la escalera; no es necesario verificar si también están presentes otros acontecimientos tales como creencias e intenciones.¹² Que el subir habría acaecido, hubiera o no acaecido la creencia racionalizadora y el deseo, demuestra por cierto la irrelevancia causal y explicativa de la creencia y el deseo.¹³ Pues una conexión explicativa puede darse sólo si existe una relación de dependencia de algún tipo; quizá, la condición que explica por qué acaeció un acontecimiento debe ser necesaria al menos en las circunstancias en que acaece el evento. Adviértase, de paso, que, tal como se lo formula, éste es un argumento general completamente independiente de las temáticas de ambas explicaciones; no hace uso del hecho de que una de las explicaciones concierne a estados psicológicos y la otra a la neurofisiología, o del hecho de que el *explanandum* concierne a la acción humana o al comportamiento. Esto está claro puesto que (I) es completamente general y temáticamente neutral [*topic neutral*].

Así formulado, el principio crucial (I) es obviamente implausible, si consideramos que las condiciones causales se dan en momentos diferentes;¹⁴ por ejemplo, sea *A* nomológicamente suficiente para *E*, y *E* para *A**, donde *A* acaece antes que *E*, y *E* antes que *A**. Entonces *A** es nomológicamente necesaria para *E*. Pero, aparentemente, aquí no hay ninguna incoherencia. O supongamos que *A* es suficiente para *A**, que acaece después, y que *A**, a su vez, es necesaria para *E*, donde *E* es posterior a *A**. Nuevamente, no hay una incoherencia evidente. Siguiendo a Goldman, en consecuencia, podríamos revisar el principio de la siguiente manera:

(II) Si *A* es suficiente para un acontecimiento posterior *E*, no es necesario para *E* ningún evento que acaezca en el mismo momento que *A* y que sea completamente distinto de él.

¹² Esto supone, como Malcolm bien sabe, que las creencias y deseos no son idénticos a los acontecimientos fisiológicos; es decir, que la así denominada tesis de la identidad acerca de lo mental es falsa. Si se da la identidad, no tenemos dos explicaciones *independientes*. Luego veremos más sobre esto.

¹³ A menos, quizá, que el comportamiento esté sobredeterminado por el acontecimiento neurofisiológico y el par deseo-creencia. Esta posibilidad se discute luego.

¹⁴ Goldman, *A Theory of Human Action*, pp. 160-61.

Y es este principio más débil el que Goldman trata de socavar en su discusión del argumento de la incompatibilidad de Malcolm.¹⁵

La objeción de Goldman a (II) es la siguiente: supongamos que dos eventos, A y A^* , son "equivalentes nómicos simultáneos"¹⁶ en el sentido de que, por razones legales, A acaece (a un objeto) en un momento determinado si y sólo si A^* acaece en ese mismo momento. Entonces, si A es suficiente para E , A^* también es suficiente para E . Si A es necesario para E también lo es A^* , y podemos suponer que A es tanto necesario como suficiente para el evento subsiguiente E . Esto significa que dos eventos distintos, A y A^* , son tales que A es suficiente y A^* es necesario para que acaezca E .

Aquí hay varios problemas complejos que involucran las interrelaciones entre necesidad, suficiencia, causa, explicación y otros conceptos similares. Pero, afortunadamente, podemos ignorarlas en gran medida, pues lo que es crucial para el problema de la compatibilidad explicativa es simplemente esta afirmación: *si A y A^* son equivalentes nómicos simultáneos, en el sentido explicado, entonces uno de ellos constituye una explicación para un evento dado si y sólo si el otro también lo es*. Si esta pretensión es correcta, la existencia de correlatos fisiológicos para las creencias y deseos debería garantizar la posibilidad de que exista tanto una explicación racionalizadora como una fisiológica de una determinada acción. Si todos los eventos mentales tienen equivalentes nómicos en estados y procesos fisiológicos (podríamos denominar a esto "la tesis de correlación psicofísica"), toda explicación racionalizadora tendría que tener una contraparte fisiológica con el mismo *explanandum*.

¹⁵ Malcolm estaría mal aconsejado si dejara descansar su argumento en este principio revisado (aunque el principio fuera válido). Pero no tendría poder para mostrar la incompatibilidad entre una explicación racionalizadora y una explicación fisiológica que hiciera uso de las condiciones fisiológicas iniciales que acaecen un poco después o antes que la creencia y los deseos invocados en la explicación racionalizadora. Lo que es crucial no es que las dos condiciones para E acaezcan al mismo tiempo; sino antes bien que pertenezcan a cadenas causales (o cadenas de condiciones) independientes y distintas.

¹⁶ De manera que estamos tratando A y A^* , y también E , como "acontecimientos tipo" o "acontecimientos genéricos". De hecho, hablar de "necesidad" y "suficiencia" parece tener sentido claro sólo para los acontecimientos genéricos. Pero puede considerarse que nuestra discusión concierne a los acontecimientos individuales si tomamos a estas últimas como instanciaciones de acontecimientos genéricos. Para detalles acerca de esta concepción de los eventos véase mi "Events as Property Exemplifications" en *Action Theory* de Myles Brand y Douglas Walton (comps.) (Dordrecht: Reidel, 1976). Sin embargo, en la discusión presente, no se presupone ninguna concepción particular sobre la naturaleza de los acontecimientos.

2. Equivalentes nómicos y explicaciones dependientes

La tesis recién mencionada que afirma que si A y A^* son equivalentes nómicos simultáneos, entonces, uno es una explicación de un acontecimiento dado sólo en el caso de que el otro también lo sea, se apoya en la teoría hempeliana de la explicación. Pero esta tesis irrestricta es sin duda sospechosa; podría ocurrir que aunque la situación fuera tal como se la describió, A^* sea sólo un “epifenómeno” de A , y aunque A^* es nomológicamente suficiente (y quizá también necesario) para E , no explica por qué acaece E .¹⁷

Por tanto, A podría ser el estado patológico subyacente de alguna enfermedad, A^* un síntoma simultáneo de este estado y E un estado posterior de la enfermedad. Un caso como éste, en consecuencia, no es uno en el que hay dos explicaciones para un único *explanandum*, pues los epifenómenos no explican.

Pero no nos extendamos en esta posibilidad (aunque, como veremos, prefigura otras que serán consideradas luego). Supongamos, en cambio, algo parecido a la teoría nomológica de suficiencia que ofrece Hempel [*hempelian nomic sufficiency account*] como nuestro modelo de trabajo de explicación (esto es lo que hacen tanto Malcolm como Goldman). El compatibilista explicativo puede estar dispuesto a aceptar la posibilidad de los epifenómenos recién mencionados, pues todo lo que necesita para refutar la pretensión de Malcolm es que se dé un caso en el que haya dos equivalentes nómicos simultáneos y en donde cada uno constituya un *explanans* para el mismo acontecimiento; por cierto, el compatibilista puede argüir que no todo caso en el que tenemos equivalentes nómicos simultáneos es tal que uno de ellos es un epifenómeno del otro. De todos modos, de acuerdo con el modelo hempeliano, si A y A^* son equivalentes nómicos, las dos explicaciones que hacen uso de ellos son “explicaciones nómicamente equivalentes” en un sentido directo; también, bajo la explicación nómica subsuntiva de las relaciones causales, A y A^* pueden ser denominadas “causas nómicamente equivalentes”. Pues el que una de ellas sea una explicación de un acontecimiento dado si y sólo si la otra también lo es y que A sea una causa de un determinado evento si y sólo si A^* es también su causa será una cuestión legal. Por tanto, si las creencias y deseos tienen coextensiones nomológicas en los estados fisiológicos, para cada explicación racionalizadora de una acción hay una explicación fisiológica nomológicamente equivalente (que podríamos formular si tu-

¹⁷ Acerca de los epifenómenos, véase, por ej., David Lewis, “Causation”, *Journal of Philosophy* 70 (1973): 556-67.

viéramos suficiente conocimiento de las leyes de correlación psicofisiológicas). El hecho de que estas explicaciones sean nómicamente equivalentes en este sentido debería alertarnos con respecto a la posibilidad de que no tengamos aquí dos explicaciones *independientes* de la misma acción. Como dije recién, el que una sea una explicación de esa acción sólo en el caso en que la otra lo sea, es una cuestión de necesidad nomológica.

Por tanto, en el tipo de situación que Goldman nos pide que consideremos, nuestra conclusión provisional parece ser la siguiente: o bien uno de los dos estados nómicamente equivalentes es un epifenómeno del otro de manera que no tenemos dos explicaciones del mismo acontecimiento, o bien tenemos dos explicaciones que no son nomológicamente independientes. Pero, ¿qué hay de malo en las explicaciones nomológicamente dependientes o equivalentes en este sentido? ¿No son suficientes para mostrar la posibilidad de ofrecer dos explicaciones distintas de uno y el mismo acontecimiento? Y si vale la tesis de la correlación psicofísica, ¿no se nos asegura la posibilidad general de explicar el comportamiento tanto racional como fisiológicamente?

Y parece haber instancias precisamente de este tipo en otras áreas de la ciencia. Por ejemplo, por una parte, podríamos explicar el comportamiento de alguna sustancia sujeta a ciertas condiciones apelando a sus propiedades físicas disposicionales más gruesas (ductilidad, conductividad, viscosidad, etc.) y, por otra parte, formulando una explicación más teórica que invoque las microestructuras que subyacen a estas disposiciones como sus “equivalentes nómicos”. ¿No tenemos en tales casos exactamente el tipo de ejemplo que se ajusta al argumento de Goldman? Tales ejemplos parecen abundar; a menudo profundizamos y enriquecemos nuestra comprensión de los fenómenos naturales alejándonos de sus rasgos observables y de las toscas “leyes fenoménicas” que los gobiernan, en dirección a sus microestructuras subyacentes e invocando “leyes teóricas” más sistemáticas apropiadas a estos estados. (Al menos, ésa es la explicación de libro de texto del progreso en la teorización científica.) Quizá, las explicaciones racionalizadoras se relacionan con las explicaciones fisiológicas justamente de esta manera —es decir, como lo hacen las macroexplicaciones con las microexplicaciones subyacentes— a través de un difundido sistema de leyes de correlación que proporciona un “equivalente nómico simultáneo” para cada estado fisiológico.

Que tales leyes de correlación existan, especialmente para los estados intencionales con contenido [*contentful*] (“actitudes proposicionales”) tales como la creencia, el deseo y la intención, es una cuestión controversial sobre la que se ha escrito mucho en las dos últimas décadas. Pienso que

hoy en día existe una preponderancia de argumentos filosóficos [*philosophical evidence*] del lado del “anomicismo psicofísico” [*psychophysical anomalism*], es decir, la tesis de que no hay, y no puede haber, leyes precisas que conecten estados intencionales con estados fisiológicos del cerebro (o cualquier estado físico).¹⁸ Si no tenemos confianza en la existencia de tales leyes de correlación, sean dichas correlaciones de carácter específico para una especie o para un tipo de estructura, o sean uniformes para todos los organismos y estructuras, la solución en términos de “equivalentes nómicos” sólo sería una posibilidad ociosa. Pero aquí no quiero proseguir con el tema de las leyes psicofísicas, pues sigue planteada la pregunta acerca de si la existencia de estados nomológicamente equivalentes nos abre o no la posibilidad de tener múltiples explicaciones de un único acontecimiento. Es decir, queremos saber cómo serían las cosas si hubiera correlaciones psicofísicas.

Lo que quiero afirmar es esto: el tipo de situación que describe Goldman, a saber, una en la que dos acontecimientos *A* y *A** son vistos como nomológicamente necesarios y suficientes uno del otro, y en la cual se piensa que cualquiera de ellos constituye un *explanans* para uno y el mismo acontecimiento *E*, constituye una situación intrínsecamente inestable. Ello es así especialmente cuando *A* y *A** son cada uno de ellos miembros de un sistema de acontecimientos (o conceptos) tal que los dos sistemas a los que cada uno respectivamente pertenece muestran el tipo de conexiones nomológicas sistemáticas que Goldman visualiza para lo psicológico y lo fisiológico. La inestabilidad de la situación genera una fuerte presión por encontrar una explicación aceptable de la relación entre *A* y *A**, por extensión, de la que se da entre los dos sistemas a los que pertenecen; la inestabilidad se disipa y se restablece el equilibrio cognitivo cuando llegamos a ver una relación más específica entre ambas explicaciones. Como veremos, en los casos interesantes, la relación específica que reemplaza a la equivalencia será la identidad o bien alguna relación de dependencia asimétrica.

Otra manera de expresar mi tesis es la siguiente: existe cierta inestabilidad en una situación en la que se pretende que dos acontecimientos distintos son causas o explicaciones nomológicamente *equivalentes* del mismo fenómeno; la estabilidad se restablece cuando *la equivalencia es reem-*

¹⁸ Quizá la argumentación más influyente para el anomicismo psicofísico es la defendida por Donald Davidson en “Mental Events”, impreso en su *Essays on Actions and Events* (New York: Oxford University Press, 1980). Véase mi “Psychophysical Laws” en *Actions and Events: Perspectives on the Philosophy of Donald Davidson* de Ernest LePore y Brian McLaughlin (comp.) (Oxford: Basil Blackwell, 1985), para una exégesis y discusión del argumento de Davidson y referencias adicionales.

plazada por la identidad o alguna relación asimétrica de dependencia. Es decir, o bien dos explicaciones (o causas) de hecho se funden en una, o bien si en verdad estamos ante dos explicaciones (o causas) distintas, tenemos que ver a una de ellas como dependiente, o derivada, de la otra; o lo que es igual, a una de ellas como adquiriendo dominancia explicativa o causal sobre la otra.

La tensión en esta situación de la que surge la inestabilidad puede ser vista de diversas maneras. En primer lugar, si A y A^* son cada uno una causa suficiente del acontecimiento E , ¿por qué entonces E no está *sobredeterminado*? En el mejor de los casos, es extremadamente raro pensar que todas y cada una de las acciones que realizamos está sobredeterminada en virtud de tener dos causas suficientes distintas.¹⁹ Por cierto, esto difiere del caso estándar de sobredeterminación en el que las dos causas sobredeterminantes no están nomológicamente conectadas. Pero ¿por qué la supuesta relación nomológica entre A y A^* convierte en vacua la pretensión de que éste es un caso de sobredeterminación causal? Advuértase la interrelación que se da aquí: cuanto más cerca nos encontramos del caso estándar de sobredeterminación, menos dependientes son ambas explicaciones entre sí y, correlativamente, cuanto más se enfatiza el punto de que no estamos frente a un caso de sobredeterminación estándar en virtud de la equivalencia nómica entre las explicaciones, menos plausible es nuestra pretensión de que nos hallamos ante dos explicaciones distintas e independientes.

En segundo lugar, si A y A^* son equivalentes nómicos, co-ocurren por razones legales —es decir, es nómicamente imposible que acaezca una sin la otra. ¿Por qué entonces no forman una causa única conjuntamente suficiente para E en vez de dos causas individualmente suficientes? ¿Cómo sabemos que cada A y A^* no es sólo una causa parcial de E ? Esto es, ¿por qué no tendríamos que considerar a A y A^* como formando una explicación única y completa de E en vez de dos explicaciones separadamente suficientes de él? ¿Cómo habremos de decidir entre ambas alternativas?

Cuando reflexionamos sobre el caso especial de la causación psicofísica, donde A , digamos, es un acontecimiento psicológico, A^* es su correlato fisiológico y E es algún movimiento corporal asociado con una acción, sería altamente implausible considerar a A como actuando directamente sobre el cuerpo para originar E (imagínese, por ejemplo, que mi creencia y mi deseo actúan telequinéticamente sobre los músculos en mi brazo y mis hombros y los hacen contraerse, causando por ello que mi brazo se

¹⁹ Goldman es consciente de este punto pero no sigue sus implicaciones; de hecho, tal como Malcolm, considera la posibilidad de que las creencias y los deseos sean estados neurales. Véase más adelante la "solución basada en la identidad".

alce); sería más creíble pensar que si el par deseo-creencia ha de causar el movimiento de mi brazo, debe “funcionar a través” de la cadena causal física que parte de A^* , pasa por algún acontecimiento neuronal en el cerebro y culmina en una contracción muscular. Si eso es correcto, no podemos considerar a A y A^* como constituyendo explicaciones independientes de E . Debemos pensar que la eficacia causal de A para originar E es dependiente de la que posee su correlato físico A^* .²⁰

Creo que estas perplejidades se disipan sólo cuando tenemos una explicación de la relación entre A y A^* , es decir, entre las dos supuestas causas de una acción única, y que, tal como argüiré, una explicación que es adecuada para esta tarea mostrará que A y A^* no pueden cada una por sí misma constituir una explicación completa e independiente de la acción.

Un caso que ilustra muy bien esto es la solución basada en la identidad: al decir que A y A^* son de hecho uno y el mismo acontecimiento, podemos resolver la situación de manera elegante. Malcolm es consciente de esto (del mismo modo que Goldman), pues en el curso de su argumentación sobre la incompatibilidad rechaza explícitamente la tesis de la identidad psicofísica.²¹ De acuerdo con la posición que afirma la identidad, tenemos una sola causa de E y no dos cuyas relaciones mutuas necesitamos explicar. Con respecto a la explicación, al menos en un sentido objetivo, aquí tenemos una sola explicación, y no dos. Ambas explicaciones difieren sólo en el aparato lingüístico que se usa para referir, o seleccionar, las condiciones y acontecimientos que realizan la tarea explicativa; son sólo variantes descriptivas una de la otra. Quizás ofrecen información causal acerca de E de maneras diferentes, cada una apropiada en un contexto explicativo particular; pero ambas señalan una única conexión causal objetiva y se fundamentan en este único hecho causal.²²

¿Qué otras explicaciones posibles de la relación entre A y A^* hay? El modelo estándar de teoría de la reducción sostiene que para que una teoría

²⁰ Creo que esta imagen puede generalizarse, véase mi “Epiphenomenal and Supervenient Causation”, *Midwest Studies in Philosophy* 9 (1984): 257-70.

²¹ Malcolm, “The Conceivability of Mechanism”, p. 134.

²² Una explicación de la identidad en el contexto presente sería una forma de la denominada teoría “de la identidad tipo” (hablar de “equivalencia nómica” entre A y A^* , por ejemplo, implica que éstos representan “acontecimientos genéricos” o acontecimientos tipo, no acontecimientos concretos); sin embargo, el punto se aplica también a la teoría de la “identidad de ejemplares”. Así pues, la teoría causal de la acción de Donald Davidson es un ejemplo: las razones (por ej., creencias y deseos) son causas de la acción, pero son redescribibles en términos físicos (presumiblemente fisiológicos) y, por lo tanto, *son* acontecimientos físicos, y es bajo sus descripciones físicas como las razones y acciones son subsumidas bajo leyes. Por lo tanto, para Davidson la dualidad de la explicación desaparece (empero, que éste sea un resultado pretendido por Davidson es otra cuestión).

sea “reducida a” otra, los predicados teóricos primitivos de la teoría a ser reducida deben ser conectados *via* “leyes puente” con los predicados, presumiblemente complejos, de la teoría base a la cual se la reduce, de manera tal de permitir la derivación de las leyes de la teoría a reducir de aquellas de la teoría reductora.²³ Está claro que si se dispusiera de una ley puente de forma bicondicional para cada predicado primitivo de la teoría a ser reducida, su reducibilidad está asegurada. Pues todo lo que necesitamos hacer es reformular las leyes básicas de la teoría a ser reducida en el vocabulario de la teoría reductora mediante el uso de las leyes bicondicionales y agregar estas reformulaciones, como es requerido, a los axiomas de la teoría reductora. De cualquier modo, el punto es que si cada evento psicológico tiene un equivalente nómico simultáneo, se satisfacen todas las condiciones necesarias para la reducción de la psicología a la fisiología, en el sentido actualmente estándar de reducción. Nuestra psicología intencional, con todas sus racionalizaciones, estaría lista para la absorción reductiva dentro de la fisiología y las ciencias asociadas y las explicaciones racionalizadoras dejarían de ser explicaciones independientes de las acciones.²⁴ La relación entre una explicación racionalizadora y la explicación fisiológica a la cual se la reduce sería, por cierto, semejante a aquella entre una macroexplicación —recurriendo una vez más el ejemplo trillado— de algún fenómeno térmico (la expansión de un gas cuando se lo calienta) y su microexplicación subyacente (en términos de las crecientes energías cinéticas de las moléculas de gas). Aquí no tratamos con dos narraciones independientes acerca del mismo fenómeno; la diferencia principal entre ellas es que una narra una historia más detallada, más reveladora y teóricamente más fecunda que la otra. Ya no existe el tipo de tensión que Malcolm trata de explotar cuando tenemos tanto una explicación racionalizadora como una fisiológica de una acción. La razón es que ambas explicaciones ya no son independientes: una es reducible a la otra.

²³ Véase Ernest Nagel, *The Structure of Science* (New York: Harcourt, 1961).

²⁴ Por lo tanto, en nuestra presente interpretación, tanto Malcolm como Goldman están en lo correcto acerca de ambos tipos de explicación. Pues nosotros interpretamos la posición de Malcolm como diciendo que las explicaciones racionalizadoras y mecanicistas no son compatibles en tanto explicaciones *independientes*, y Goldman *no* pretende que ambos tipos de explicaciones sean independientes en nuestro sentido. En “Mechanism and Responsibility”, Daniel C. Dennett también toma el tema de la compatibilidad; sin embargo, su foco es diferente. Su interés principal es mostrar que el comportamiento de los mecanismos finitos, “sistemas tropísticos”, puede ser explicado a partir de “un enfoque intencional” [*intentional stance*] aun cuando se acepte la conclusión de Dennett, nuestro problema sigue estando en pie: ¿cuál es la relación entre “las explicaciones de enfoque intencional” y las “explicaciones de enfoque físico”? ¿Son compatibles ambos tipos de explicación cuando se da el mismo comportamiento?

En consecuencia, la supuesta existencia de coextensiones nómicas psicofísicas no muestra que las explicaciones racionalizadoras y fisiológicas puedan coexistir como explicaciones independientes de las acciones; por el contrario, mostraría que las explicaciones de un tipo son reductivamente dependientes de las del otro tipo. Pues nos colocaría precisamente en una situación hecha a medida para la reducción fisiológica de la psicología, a saber, una en la que la explicación racionalizadora se verá privada de sus estatus de modo independiente de comprender las acciones, una situación que, como vimos, Malcolm quería excluir. Si se lograra tal reducibilidad, quedaría socavada la pretensión de que la explicación racionalizadora nos proporciona un modo distintivo de comprensión de la acción humana. La eficacia explicativa que posee la explicación racionalizadora derivaría de la de la explicación fisiológica subyacente, al menos en este sentido: estaríamos en condiciones de ofrecer una explicación de por qué y cómo es que las razones explican la acción y, si adoptamos una concepción causal de la situación, entonces precisamente de cómo las razones causan las acciones y mediante qué mecanismo.

3. Explicación causal y exclusión explicativa

El principio general de exclusión explicativa afirma que no pueden coexistir dos o más explicaciones completas e independientes del mismo acontecimiento. Los significados de “completas” e “independientes” son obviamente cruciales. No ofreceré definiciones de estos términos; antes bien, me centraré en algunos casos específicos que caen bajo las distinciones prometidas, con la esperanza de que, en el curso de mi discusión surgirán núcleos de significado razonablemente determinados que brindarán al principio de exclusión un contenido lo suficientemente claro y sustancial. Un examen minucioso de la exclusión explicativa inevitablemente se extenderá hasta abarcar el viejo debate acerca de la naturaleza de la explicación, un tema sobre el que en la actualidad no existe nada parecido al consenso. La discusión que sigue inevitablemente se apoyará en ciertas suposiciones intuitivas acerca de cómo operan las explicaciones, en especial las explicaciones causales. 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 ficcional o instrumentalista de la explicación, el principio de exclusión explicativa es una restricción general plausible.

Me parece que el argumento de la exclusión explicativa se muestra más persuasivamente con respecto a las explicaciones causales de acontecimientos individuales. Supongamos, pues, que tenemos dos explicacio-

nes causales de un acontecimiento simple:

La explicación A menciona a A como una causa de E .

La explicación B menciona a A^* como una causa de E .

¿Qué hemos de pensar de tal situación? Se pueden distinguir diversas posibilidades:

Caso 1. Encontramos que $A=A^*$. Es decir, que aquí hay una causa, no dos. Vimos cómo opera esto en el caso de la causación psicofísica bajo la teoría de la identidad mente-cuerpo.²⁵ Este caso, cuando se da, provee la manera más simple y quizá la más satisfactoria de aliviar la tensión creada por la existencia de ambas explicaciones. Tales identidades a menudo se encuentran en un contexto reductivo donde una de las explicaciones especifica la causa mediante una descripción más profunda y más teórica y sistemática.

Caso 2. A es distinta de A^* , pero en algún sentido claro es “reducible a” o “superveniente” de A^* . Este tipo de situación surgirá en un contexto reductivo como el que recién consideramos, siempre y cuando por alguna razón nos abstengamos de *identificar* los acontecimientos o estados reductivamente relacionados. Por tanto, el caso psicofísico considerado antes sería una instancia de este tipo si creyéramos en la reducibilidad nomológica de los estados psicológicos a los estados físicos, sin no obstante desear identificar los estados psicológicos con sus correlatos físicos; o si creyéramos en la “supervenencia” en un sentido apropiado de lo psicológico sobre lo físico, sin identificar un estado psicológico superveniente con su base física. En tal situación, es posible tratar a las relaciones causales que involucran acontecimientos y estados psicológicos como siendo ellas mismas supervenientes de procesos causales físicos más fundamentales o reducibles a ellos. He discutido extensamente la “causación superveniente” en otra parte.²⁶ De cualquier modo, no tenemos en casos de este tipo dos explicaciones causales *independientes* del mismo acontecimiento. Ambas explicaciones pueden coexistir porque una de ellas depende de la otra, de manera reductiva o por supervenencia.

²⁵ La formulación precisa aquí se vería afectada en alguna medida según si tomamos la teoría de la “identidad de ejemplares” o la de “identidad de tipo” y de acuerdo con la concepción acerca de la naturaleza de los acontecimientos que se adopte. Sin embargo, el punto general debería aplicarse sean cuales fueren las posiciones que se adoptaran acerca de estas cuestiones.

²⁶ En “Causality, Identity, and Supervenience in the Mind-Body Problem”, *Midwest Studies in Philosophy* 4 (1979): 31-49; y “Epiphenomenal and Supervenient Causation”, *Midwest Studies in Philosophy* 9 (1984): 257-70. Véase también Ernest Sosa, “Mind-Body Interaction and Supervenient Causation”, *Midwest Studies in Philosophy* 9 (1984): 271-81. Para una discusión general de la supervenencia véase mi “Concepts of Supervenience”, *Philosophy and Phenomenological Research* 65 (1984): 153-76.

Caso 3. Ni A ni A^* son en sí mismas una “causa suficiente” de E aunque cada una es un componente indispensable de una causa suficiente. Como se ha observado a menudo, se nos pide que proporcionemos una causa o una explicación para un acontecimiento que usualmente seleccionamos como un factor causal que, por diversas razones epistemológicas o pragmáticas, se cree que es la más apropiada para la situación. Un ejemplo típico puede expresarse como sigue: podríamos explicar por qué ocurrió un accidente automovilístico citando, digamos, el tránsito congestionado, la ruta con hielo, los frenos en malas condiciones o la inexperiencia del conductor, etc., dependiendo del contexto explicativo, cuándo cada una de estas condiciones desempeñará un papel esencial al causar el accidente. Si A y A^* están relacionados de esta manera no tenemos dos explicaciones *completas*, al menos en un sentido de “explicación completa”, a saber, uno en el que una explicación completa especifica un conjunto suficiente de condiciones causales para el *explanandum*. Está claro que pueden coexistir dos explicaciones incompletas, en este sentido, del mismo acontecimiento. El principio de exclusión explicativa sólo prohíbe que haya más de una explicación completa e independiente del mismo acontecimiento.

Caso 3a.²⁷ A es parte propia de A^* . Si esto es así, A , en tanto explicación de E , no es en sí misma ni completa ni independiente de A^* .

Caso 4. A y A^* son eslabones diferentes en la misma cadena causal que conduce, digamos, de A a A^* y luego a E . En este caso nuevamente no tenemos dos explicaciones causales independientes; el *explanans* de uno, A^* , depende causalmente del *explanans* del otro, A .

Caso 5. A y A^* son distintos y cada uno es una causa suficiente de E . Podemos pensarlos como perteneciendo a dos cadenas causales distintas e independientes. Este es entonces un caso de *sobredeterminación causal*: E habría acaecido aun cuando algunos de los dos A y A^* no hubiesen acaecido, o no hubieran sido sus causas; el otro habría sido suficiente para originarlo. Así, un hombre es herido de muerte por dos asesinos cuyas balas le pegan al mismo tiempo; o un edificio se incendia por un corto circuito en la instalación defectuosa y un rayo que cae en él, en el mismo instante. En casos como éste no es obvio cómo tendríamos que formular una explicación de por qué o de qué manera se originó el acontecimiento sobredeterminado. Sin embargo, no es implausible pensar que el no mencionar cualquiera de las causas sobredeterminantes proporciona una imagen engañosa e incompleta de lo que ocurrió, y que ambas causas tendrían que figurar en cualquier explicación *completa* del evento. Si ello es

²⁷ Le debo este caso a Karl Pfeifer.

correcto, éste no es un caso en el que son posibles dos explicaciones completas e independientes del mismo acontecimiento.

Por tanto, no concuerdo con Hempel cuando dice que cuando tenemos un caso de "sobredeterminación explicativa", en el que hay dos o más argumentos N-D completos con el mismo enunciado *explanandum* (por ejemplo, "La longitud de esta barra metálica se incrementó"), entonces cualquiera de los argumentos puede *individualmente* ser considerado "completo" (uno de ellos podría explicarlo invocando que la barra fue calentada y otro podría explicarlo mencionando el hecho de que se la sometió a estiramiento longitudinal).²⁸ Dado el interés dominante que Hempel tiene en la dimensión predictivo-inferencial de la explicación, su posición al respecto no es sorprendente. Pues en tanto inferencia predictiva cualquiera de los dos argumentos N-D es completo y suficiente. Sin embargo, cuando lo que buscamos es una explicación causal del alargamiento de la barra, cuando queremos saber *por qué* se incrementó la longitud de la barra, la situación parece alterarse radicalmente; nuestra comprensión de por qué acaeció el evento es en el mejor de los casos incompleta, y quizá defectuosa, si no tenemos conciencia de alguna de las dos "explicaciones".

El sentido de "completitud" de una explicación que acabo de invocar es diferente del que usara al caracterizar el caso de "causa parcial" (caso 3 antecedente), aunque esto no impide que haya un sentido más amplio que incluya a ambos. No obstante, no es crucial para mis tesis generales acerca de la exclusión explicativa conocer la manera exacta en que tratamos los casos de sobredeterminación explicativa, pues es improbable que aquellos que quieran admitir las explicaciones múltiples de un acontecimiento único deseen restringirse a las instancias de sobredeterminación causal. El principio de exclusión conservaría el contenido sustancial aun cuando se dejaran al margen los casos de sobredeterminación. De todos modos, el punto importante a tener en claro es el siguiente: que tengamos en manos un caso de sobredeterminación causal es una manera en la que puede funcionar una explicación satisfactoria de la relación entre *A* y *A**, eliminando las perplejidades generadas por la pretensión de que cada una es causa de algún acontecimiento único.

Estas consideraciones sugieren un argumento simple en favor de la exclusión explicativa para las explicaciones causales: supongamos que se alega que *A* y *A** ofrecen cada uno una explicación completa de *E*. Consideremos las dos preguntas siguientes: 1) ¿habría acaecido *E* si *A* no hu-

²⁸ Carl G. Hempel, *Aspects of Scientific Explanation* (New York: The Free Press, 1965), p. 419.

biera acaecido? y 2) ¿habría acaecido E si A^* no hubiera acaecido? Si la respuesta a ambas preguntas es “sí”, éste es un caso clásico de sobre-determinación y, como se señaló en los párrafos anteriores, podemos tratar este caso como uno en el que cualquier explicación tomada de manera aislada es incompleta, en caso contrario, exceptuarla de la exclusión explicativa. Si la respuesta a al menos una de ellas, digamos la primera, es “no”, debe ser porque si A no hubiera acaecido, A^* tampoco lo habría hecho. Y esto significa que A y A^* no son independientes, y por tanto, que ninguna de las dos explicaciones es una explicación independiente de E .

La discusión previa acerca de los subcasos es útil como manera de hacer más claro qué puede querer significarse con “completitud” e “interdependencia” de explicaciones. Cuando examinamos las posibilidades particulares que parecen permitir dos explicaciones distintas de un acontecimiento, parecemos ser capaces de encontrar —y nos vemos compelidos a buscar— razones para decir que no son independientes o al menos que una de ellas no es completa. Dos explicaciones de un acontecimiento crean cierta tensión epistémica, una tensión que se ve disipada sólo cuando tenemos una explicación de cómo ellas, o las dos causas por ellas indicadas, se relacionan entre sí. Lo que necesitamos para aliviar la tensión es descubrir cuál de las causas examinadas vale para el caso dado.

4. Observaciones acerca de la epistemología y de la metafísica de la explicación

Cuando buscamos una explicación de un acontecimiento típicamente estamos en un estado de perplejidad, es decir, en un tipo de encrucijada epistémica [*epistemic predicament*].²⁹ Una explicación exitosa nos sacará de este estado. Si nuestra discusión no ha estado mal encaminada, lo que muestra es que un exceso de explicaciones nos volverá a conducir a una encrucijada epistémica similar, la que sólo se verá aliviada cuando tengamos una explicación acerca de la manera como se relacionan las explicaciones entre sí.

Quizás encontremos en lo que sigue una explicación de por qué esto es así. Algunos escritores han enfatizado el rol unificador y simplificador de la explicación y han tratado de conectarlo con la comprensión.³⁰ Tiene sentido pensar que explicaciones múltiples de un *explanandum* único son

²⁹ Sylvain Bromberger lo denomina “*p-predicament*” o “*b-predicament*” en su “An Approach to Explanation” en *Analytic Philosophy*, de R. J. Butler (comp.), 2nd series (Oxford: Basil Blackwell, 1965).

³⁰ Véase, por ejemplo, Michael Friedmann, “Explanation and Scientific Understanding”, *Journal of Philosophy* 71 (1974): 5-19; Philip Kitcher, “Explanation, Conjunction, and Unification”, *Journal of Philosophy* 73 (1976): 207-212.

presumiblemente contraproducentes en relación con el objetivo de simplificación y unificación. Cuando se producen dos explicaciones distintas para dar cuenta de un fenómeno único, parecemos encaminarnos en una dirección exactamente opuesta a la máxima de simplificación explicativa: "Explique tanto como pueda con la menor cantidad de premisas explicativas". La unidad se logra buscando interconexiones entre los ítems del conocimiento y la simplicidad aumenta cuando estas interconexiones son vistas o interpretadas como relaciones de dependencia. Pues el rol principal de las relaciones de dependencia en un sistema es que ayudan a reducir el número de suposiciones independientes o primitivas requerido. Si la simplicidad y la unidad de la teoría constituyen nuestra meta cuando buscamos explicaciones, las explicaciones múltiples de un fenómeno único son contraproducentes, a menos que seamos capaces de determinar que sus premisas explicativas están relacionadas entre sí de maneras apropiadas. Queda claro que mostrar que las premisas explicativas de una explicación dependen de las de la otra, en cualquiera de los sentidos de dependencia distinguidos en la sección precedente en conexión con las explicaciones causales, es, en efecto, un intento de reducir el número de premisas explicativas independientes: en otras palabras, constituye un movimiento en dirección a restablecer el rol simplificador y unificador de las explicaciones.

Estas reflexiones, aun cuando sean esquemáticas y programáticas, nos proporcionan una clave para lograr una manera posible de comprender el concepto de dependencia para las explicaciones, una noción de la que hemos hecho uso de un modo bastante liberal sin una explicación general. Un tema que cruza los diversos casos de "explicaciones dependientes" que hemos expuesto parece ser el que sigue: si una explicación depende de otra, ambas explicaciones tomadas en conjunto no están comprometidas con más suposiciones independientes acerca del mundo de lo que lo está la segunda explicación tomada separadamente. Esto encaja muy bien con la concepción según la cual las explicaciones aumentan la comprensión mediante la simplificación y la unificación. En este caso, estaríamos cambiando la noción de "explicación independiente" por la de "suposición independiente"; pero pienso que esto mismo puede constituir un progreso.

Estas son consideraciones generales que no se restringen a las explicaciones causales. Sin embargo, la exclusión explicativa parece más obvia, y casi trivial, para la explicación causal de los acontecimientos individuales, y la razón parece ser ésta: pensamos a estas explicaciones como directamente ligadas a las historias causales efectivas de los acontecimientos que

se están explicando, y su "corrección" en tanto explicaciones está determinada por la exactitud con que retratan las conexiones causales tal como existen. Cuando se nos presentan dos explicaciones causales de un evento, necesitamos saber cómo están situadas una en relación con la otra en el mapa causal del acontecimiento; lo que no se puede permitir que ocurra, si nuestras explicaciones han de ser coherentes, es que ambas nos cuenten dos historias diferentes acerca de la misma región del mapa causal.³¹ La exclusión explicativa puede parecer obvia para las explicaciones causales, pero ello no equivale a decir que sea trivial; las consideraciones previas suponen al menos el "realismo causal", es decir, la creencia de que existe un hecho objetivo determinado correspondiente a la historia causal de cualquier acontecimiento dado.³² También tenemos que suponer ciertos principios metafísicos acerca de la causación, por ejemplo, el principio (II) citado en la sección 1.

Considero que *explicar* es una actividad epistemológica y que una *explicación* en el sentido de "producto" o "contenido teórico" de tal actividad,³³ es algo acerca de lo cual tenemos diversas actitudes cognitivas (por ejemplo, aceptar, dudar, tener pruebas en favor, etc.). Necesitar una explicación es estar en un estado epistémicamente incompleto e imperfecto y obtener una explicación es mejorar nuestra situación epistémica; representa una ganancia epistémica. Sin embargo, el conocimiento debe concernir al mundo real: saber que *p* requiere la verdad de *p* y tener una explicación causal de un acontecimiento exige que el acontecimiento especificado como su causa sea en la realidad una causa de ese acontecimiento. Concentrémonos una vez más en las explicaciones causales de los acontecimientos individuales y dejemos a un lado otros tipos de explicación (por ejemplo, la explicación de leyes o regularidades, del significado de una palabra, de cómo interpretar una regla, de cómo opera una prueba matemática, etc.). El tipo de concepción a la que acabo de aludir, a saber, la que afirma que una explicación causal de *E* en términos de *A* es una "explicación correcta" sólo si *A* es en la realidad causa de *E*, puede denominarse una concepción "objetivista" o "realista" de la explicación. Y la concepción de que las explicaciones deben ser "reales" u "objetivas" en este sentido puede denominarse *realismo explicativo*.

³¹ Cf. la noción de "Texto explicativo ideal" de Peter Railton en "Probability, Explanation, and Information", *Synthese* 48 (1981): 233-256; también David Lewis, "Causal Explanation" en *Philosophical Papers II* (New York: Oxford University Press, 1986).

³² Discuto con algún detalle esto y temas relacionados en "Explanatory Realism, Causal realism, and explanatory Exclusion", por aparecer en *Midwest Studies in Philosophy*.

³³ Para algunas de las distinciones terminológicas básicas, véase Bromberger, "An Approach to Explanation" y Peter Achinstein, *The Nature of Explanation* (New York y Oxford: Oxford University Press, 1983), especialmente la introducción y el cap. 1.

En términos más generales, una concepción realista de la explicación sostiene que nociones tales como “verdad objetiva”, “corrección” o “exactitud” tienen sentido cuando se aplican a las explicaciones, que lo tienen de un modo más o menos literal, y que una explicación es correcta o exacta *en virtud de* que se dé “en el mundo real” cierta relación determinada entre el *explanandum* y lo que se aduce como explicación de él. Además, esta concepción realista sostiene que una explicación constituye un incremento real en el conocimiento sólo si tiene esta propiedad de corrección y exactitud, y que una explicación es epistémicamente aceptable sólo si tenemos buenas razones para pensar que la relación explicativa que pretende retratar de hecho se da en este sentido objetivo. En síntesis, tal teoría sostiene que las explicaciones se evalúan adecuadamente sobre la base de criterios objetivos tales como la exactitud y la verdad. Decir todo esto no implica menospreciar la dimensión epistemológica de la explicación. Por el contrario: así como el conocimiento de que algo tiene determinada característica queda eliminado si la cosa no posee tal característica, una explicación de *X* en términos de *Y* resulta vaciada de contenido si *Y* no existe o no se da en realidad una relación apropiada entre *X* e *Y*. Y así como la pretensión de saber que *p* es descartada, y debe serlo, cuando hay razones para creer que *p* no se da, una explicación de *X* en términos de *Y* debe ser descartada, de acuerdo con el realismo explicativo, si tenemos razones para pensar que la relación pretendida entre *X* e *Y* no se da de hecho. Cosas tales como la comprensión, la satisfacción intelectual, tornar inteligibles las cosas, resolver enigmas e inconsistencias aparentes, etc., son cruciales; sin embargo, no creemos que estas cosas constituyan *logros legítimos* [*are properly earned*] a menos que la explicación sea correcta en algún sentido objetivo. El argumento en favor del realismo explicativo se ve mejor —al menos se lo puede mostrar de manera más explícita— con respecto a las explicaciones causales, pues aquí la noción de un “correlato objetivo” de la explicación tiene un sentido aceptablemente claro, lo suficientemente claro como para una consideración explícita.³⁴

Es útil distinguir entre dos versiones del principio de exclusión. Supongamos que para algún acontecimiento dado tú tienes una explicación y yo tengo otra explicación distinta. Desde un punto de vista epistemológico puede ser racional para ti aceptar la tuya y para mí aceptar la mía.

³⁴ Algunos querían *analizar* la noción misma de causación en términos de explicación, pero creo que esto es un error. Véase, por ej., Michael Scriven, “Causation as Explanation”, *Nous* 9 (1975): 3-16; Kim, “Causes as Explanations: A Critique”, *Theory and Decision* 13 (1981): 293-309.

Pero, ¿pueden ambas ser explicaciones “correctas” o “verdaderas”? El principio *metafísico* de exclusión explicativa dice lo siguiente: ambas pueden ser explicaciones correctas sólo si al menos alguna de las dos es incompleta o una depende de la otra.³⁵ Hay un principio de exclusión *epistemológico* correspondiente: *nadie puede aceptar ambas explicaciones a menos que tenga una explicación apropiada de cómo se relacionan entre sí*. Lo que ha de valer como “explicación apropiada” de la relación, en el caso de las explicaciones causales, se encuentra ilustrado en la sección precedente; idealmente tenemos que saber cuál de estos casos se da, o al menos contentarnos con que uno de esos casos se dé, aunque no sepamos cuál de ellos.

Agregaría algunas observaciones acerca de la individuación de los *explananda*. Algunos creen que un *explanandum* queda fijado sólo cuando queda fijado un enunciado que lo representa, una visión estrechamente vinculada a una concepción inferencial de la explicación;³⁶ esto a menudo es vinculado con la concepción de que no son los acontecimientos en tanto ocurrencias concretas sino los *aspectos* de los acontecimientos (o por qué un acontecimiento dado posee ciertas propiedades) los que constituyen los objetos propios de la explicación. Así, no explicamos, digamos, que los japoneses se rindieron a las fuerzas aliadas o el accidente de Harry sino antes bien cosas tales como, por ejemplo, por qué la rendición japonesa se produjo tan tarde como lo hizo o cómo Hirohito fue capaz de superar las objeciones de los poderosos líderes militares, o por qué el auto de Harry se salió de la ruta a pleno día. Sin embargo, estas disputas parecen en gran medida irrelevantes para mis intereses: si son los aspectos de acontecimientos, en vez de los acontecimientos *simpliciter*, los que han de ser explicados, entonces la exclusión explicativa debería aplicarse a estos aspectos de acontecimientos.³⁷ Si las ocurrencias concretas son explicables, al menos en el sentido de explicar *por qué acaecieron*, entonces también ellas estarían sujetas a exclusión explicativa. Aquí las cuestiones son algo complicadas porque hay diferentes concepciones acerca de la interpretación adecuada de los acontecimientos en contextos causales y ex-

³⁵ Podría ser posible que dos explicaciones dependieran cada una de una tercera, sin que una de ellas dependiera de la otra. Parece que en tal situación ambas explicaciones podrían conservarse.

³⁶ Véase Hempel, *Aspects of Scientific Explanation*, pp. 421-22.

³⁷ Estos parecen similares a lo que Fred Dretske ha denominado “acontecimientos alo-morfos” en “Referring to Events” incluido en Peter A. French *et al.* (comps.), *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: University of Minnesota Press, 1977).

plicativos.³⁸ Pero tratar de atender estas complejidades y sutilezas sería un ejercicio en gran medida inútil de precisión filosófica para los propósitos presentes.³⁹

El principio metafísico tiene sentido sólo si se acepta alguna forma de realismo explicativo: es decir, sólo si tiene sentido, sentido literal, hablar de “corrección”, “exactitud” y “verdad” de las explicaciones. Y la aceptación del principio metafísico de exclusión explicativa nos proporcionará una base para mantenernos bajo la regla epistemológica de exclusión. Sin embargo, pienso que el principio epistemológico puede aceptarse aun cuando se rechace el realismo explicativo. Aun si renunciamos a la idea de que hay alguna relación explicativa objetiva en el mundo, podemos sin embargo encontrar algo cognitivamente perturbador y disonante en el hecho de tener que enfrentar o aceptar dos o más explicaciones independientes del mismo fenómeno. Las premisas explicativas de una explicación no precisan contradecir lógicamente a las de otra, y puede haber suficientes garantías probatorias para pensar que cada conjunto es verdadero. Sin embargo, aceptar los dos conjuntos de premisas como constituyendo explicaciones del mismo acontecimiento (o de cualquier cosa), cada uno completo en sí mismo e independiente del otro, puede inducir a un tipo de incoherencia en nuestro sistema de creencias. Esta puede ser una instancia de incoherencia epistémica que no constituye un caso de inconsistencia o incoherencia lógica (deductiva o inductiva). Antes intenté explicar por qué esto podría ser así sobre la base de la posición según la cual el rol epistémico primario de la explicación es la simplificación y la unificación de nuestro sistema de creencias. En relación con la teoría coherentista de la justificación, se menciona a menudo en forma destacada a la “coherencia explicativa” —cómo en un sistema de creencias las relaciones explicativas generan una coherencia que es mutuamente sustentadora. Si estas especulaciones tienen algún sentido, puede sintetizarse de la siguiente manera: *demasiadas explicaciones pueden ser una fuente de incoherencia en vez de ser fuente de una mayor coherencia.*

³⁸ Para una discusión de estas cuestiones y referencias adicionales véase mi “Causation, Nomic Subsumption, and the concept of Event”, *Journal of Philosophy* 70 (1973): 217-36.

³⁹ Robert Cummins ha realizado una interesante distinción entre explicación “por subsumción” (bajo una ley causal) y explicación “por análisis” (en partes componentes); véase su *The Nature of Psychological Explanation* (Cambridge: The MIT Press, 1983), esp. caps. 1 y 2. ¿Podrían darse explicaciones de estos dos tipos para uno y el mismo *explanandum*? De acuerdo con Cummins, sin embargo, la explicación subsuntiva explica *cambios* y la explicación analítica explica *propiedades*, de manera que los *explananda* son diferentes. También parece posible interpretar a las dos explicaciones como mutuamente complementarias, pero cada una siendo sólo una explicación parcial de un *explanandum* único bajo una individuación no detallada de los *explananda*.

Es interesante contrastar esta situación con las *predicciones* o *pruebas*. Tal como señalé brevemente antes, parece que a diferencia de la sobre-determinación explicativa, la sobre-determinación predictiva no crea ningún tipo de tensión epistémica y ninguna necesidad de buscar una explicación de cómo se relacionan los dos argumentos predictivos, que predicen ambos el mismo fenómeno. La sobre-determinación predictiva o inferencial es simplemente una cuestión de sobreabundancia de pruebas. Infiero que Peggy está en casa porque puedo oír la escribir a máquina y también porque las luces están encendidas. Predecimos que una barra de acero se alargará porque vemos que la estiran y sabemos que la están calentando. Aquí no hay problemas.⁴⁰ La necesidad de una explicación de sus relaciones surge cuando se invierte a tales inferencias de valor explicativo. Y ello es así, me parece, porque las explicaciones hacen afirmaciones acerca de cómo están conectadas en el mundo las cosas, afirmaciones que están ausentes en las meras pruebas o inferencias. Si esto es correcto, el hecho de la exclusión explicativa muestra que las explicaciones no pueden ser interpretadas como meras pruebas o argumentos.⁴¹

Como dije, el realismo explicativo parece adecuarse cómodamente a la exclusión explicativa aun cuando no es implicado, pienso, por ella. Una posibilidad interesante es *no* argüir en favor de la exclusión explicativa sobre la base del realismo explicativo, sino antes bien ir en dirección opuesta, a saber, ofrecer consideraciones independientes que favorezcan la versión epistemológica de la exclusión explicativa y entonces defiendan al realismo explicativo como su presentación más natural.

Antes de pasar a la sección final me gustaría considerar una posible objeción. Podría decirse que no hay necesidad de apelar a ningún aspecto epistemológico o metafísico que sea especial, y potencialmente controvertible, de la explicación para justificar algo tal como la regla de exclusión explicativa, y que todo lo que precisamos en realidad es la navaja de Ockham, es decir, el conocido principio de simplicidad, que prescribe que nos quedemos con la menor cantidad posible de entidades, hipótesis,

⁴⁰ Tampoco cuando vemos que pueden darse dos pruebas diferentes de las mismas proposiciones matemáticas. Algunos filósofos piensan, sin embargo, que algunas pruebas son más "explicativas" que otras, en el sentido de que parecen ofrecernos una explicación de "lo que hace que el teorema valga". No obstante, pienso que podría argumentarse que el mismo principio de exclusión debe aplicarse a múltiples *pruebas explicativas* de la misma proposición matemática.

⁴¹ Estas observaciones se relacionan con la discusión estándar del "realismo" vs. "instrumentalismo" acerca de las teorías científicas; en especial, la visión común de que la concepción instrumentalista de las teorías científicas, aun cuando es viable en tanto explicación de su utilidad predictiva, las priva de importancia explicativa.

principios teóricos y, por supuesto, explicaciones. En contestación, advertiría en primer término que el requisito general de simplicidad es vago y que su aplicación requiere una interpretación más precisa de la situación a la que ha de aplicarse. En particular, necesitamos determinar exactamente en qué momento las entidades en cuestión comienzan a multiplicarse “más allá de lo necesario”. De hecho, determinar dónde comienza el exceso de equipaje es la parte difícil; el resto es trivial. El principio de exclusión hace el trabajo difícil: dice que dado cualquier acontecimiento, el tener más de una explicación completa constituye un exceso de equipaje. Más aun, el principio de exclusión nos ayuda a responder la siguiente pregunta: si, como se piensa usualmente, las explicaciones representan ganancias epistémicas, ¿por qué dos explicaciones no son mejores que una? No es en absoluto obvio que las consideraciones de parsimonia por sí solas puedan ordenarnos rechazar todas las explicaciones salvo una. Por cierto podemos pensar a la exclusión explicativa como un caso especial que cae bajo el requisito general de simplicidad; es una regla específica que concierne a una manera importante en que se gana simplicidad en cuestiones explicativas y explica por qué ha de desearse esta forma de simplicidad. Es decir, que el principio de exclusión explicativa nos proporciona una razón [*a rationale*] para la aplicación de la navaja de Ockham a las explicaciones múltiples de un *explanandum* único.

5. Aplicaciones

En esta sección final quiero describir dos ejemplos extraídos de las discusiones filosóficas recientes en las que parece emplearse el principio de exclusión explicativa de manera tácita pero crucial. Los ejemplos que discutiré no pretenden constituir un argumento en favor del principio de exclusión, antes bien, pretenden mostrar que a menudo el principio es aceptado y presupuesto aunque sea de manera implícita. Alternativamente, pueden pensarse como “aplicaciones” del principio de exclusión explicativa. Pruebas como ésta muestran que, incluso al margen de consideraciones teóricas generales, el principio conlleva un grado de plausibilidad *prima facie* y no debería ser rechazado sin buenas razones.

La explicación de la reducción de teorías a la que nos referimos antes y que fuera formulada por Ernest Nagel es un modelo de *reducción conservadora*. Pues en una reducción de este tipo la teoría reducida sobrevive a la reducción siendo conservada como una subteoría de la teoría reductora. Se conservan sus conceptos mediante una vinculación, *via* “leyes puente”, con los conceptos de la teoría reductora y sus leyes se ven reen-

carnadas como leyes derivadas de una teoría más fundamental. Que una teoría sea reducida de este modo a otra cuya legitimidad no está en cuestión constituye para la primera una reivindicación o legitimación.

Hay otra explicación de la reducción, debida a John Kemeny y Paul Oppenheim,⁴² que se piensa que nos proporciona un análisis de *reducción eliminativa* o *sustitutiva*. De acuerdo con este modelo, una teoría se reduce a otra precisamente en el caso en que todos los datos explicables por parte de la primera teoría son explicables por la segunda. En los casos interesantes, la segunda, es decir, la teoría reductora, se desempeñará mejor que la primera, la teoría reducida. No hay necesidad de conexiones conceptuales o nomológicas directas entre las teorías mismas; sus vocabularios teóricos pueden ser disyuntos y puede no haber "leyes puente" que las conecten. Tampoco es preciso que entre ambas haya incompatibilidad lógica o una relación inductiva negativa o probatoria [*evidential*]. Ocurre tan sólo que una teoría, la reductora, hace el trabajo mejor que la teoría reducida (o al menos tan bien como ella), en relación a su dominio compartido.

Está claro que cualquier caso de reducción nageliana puede interpretarse como un caso de reducción a la Kemeny-Oppenheim, de modo que no es estrictamente correcto caracterizar a todos los casos de la última como reducciones sustitutivas o eliminativas. Pero también está claro que cualquier caso de reducción a la Kemeny-Oppenheim que no sea a la vez un caso de reducción a la Nagel es tal que involucra, o debería hacerlo, la eliminación de la teoría más débil reducida en favor de la teoría más rica que la reduce. Así, la teoría de la combustión basada en el flogisto fue reducida en el sentido de Kemeny-Oppenheim a la teoría de la oxidación, y fue reemplazada por ella. La teoría del *impetus* del movimiento fue reducida y eliminada en el mismo sentido por la teoría dinámica moderna del movimiento. Y lo mismo sucede en otros casos. Aquí parece estar operando un principio general como el que sigue: *si una teoría se confronta con otra más explicativa, sólo puede sobrevivir si es reducida de manera conservadora a esta última*.

La pregunta que quiero plantear es por qué se da esto, o al menos debería darse así. ¿Por qué deberíamos reemplazar, y abandonar, la teoría reducida a la Kemeny-Oppenheim en favor de su teoría reductora? Advuértase que la teoría reductora en general no *excluye lógicamente* a la teoría reducida; no es preciso que haya incompatibilidad lógica entre ambas. Además, la teoría reducida no necesita haber sido refutada; de hecho, en la medida en que haya evidencias directas puede haber buenas ra-

⁴² "On Reduction", *Philosophical Studies* 7 (1956): 6-19.

zones para pensar que es verdadera. Entonces, ¿por qué no conservar ambas? Adviértanse las consecuencias de abandonar la teoría reducida: sus entidades teóricas características, sus propiedades, acontecimientos y estados ya no son reconocidos como “reales”. Comparten el destino de la teoría descartada: el flogisto desapareció cuando la teoría del flogisto fue desechada.

Traigo a colación estas observaciones por la razón de que ayudan a dar estructura al debate actual acerca del futuro de la psicología ordinaria, la rica y heterogénea colección de perogrulladas y lugares comunes sobre los motivos, los deseos, las creencias, los anhelos, las acciones, etc. Es en términos de tales perogrulladas cómo explicamos, y predecimos al menos de manera limitada, el comportamiento de nuestros congéneres y de nosotros mismos. Nos suena inconcebible que podamos prescindir completamente de este marco de la psicología intencional; no está claro que nuestra concepción de nosotros mismos como personas y agentes pueda sobrevivir a la pérdida del esquema psicológico ordinario.

Sin embargo, muchos filósofos han planteado dudas acerca de la realidad de la psicología y de la realidad de estados tales como creencias, deseos e intenciones. La idea es que la “ciencia cognitiva”, que se desarrolla y expande con rapidez, probablemente reemplazará a la psicología ordinaria de manera tal que en algún momento del futuro será racional concluir que no hay cosas tales como deseos y creencias y que no las hubo nunca. Pero, ¿en qué sentido de “reemplazar”? ¿Cómo podría relacionarse la psicología ordinaria con la ciencia cognitiva si ha de sobrevivir y si los estados que reconoce, tales como las creencias y los deseos, han de seguir siendo reconocidos como reales?

Quienes argumentan en favor de la eliminación potencial y probable de la psicología ordinaria y de los estados psicológicos intencionales que constituyen su núcleo central, a menudo señalan dos consideraciones: primero, en comparación con la ciencia cognitiva sistemática, la ordinaria padece defectos explicativos, y segundo, no hay perspectivas de reducirla a una teoría científica sistemática. Por ejemplo, Paul Churchland, un vigoroso defensor de esta posición, escribe:⁴³

En tanto ejemplos de fenómenos mentales centrales e importantes que permanecen en gran medida o por completo en el misterio dentro del marco de la PP [psicología popular], consideramos la naturaleza y dinámica de la enfermedad mental, la facultad de la imaginación creativa... la naturaleza y funciones psicológicas del sueño... la capacidad común de atajar una pelota mientras corremos... la elaboración interna de una imagen visual tridimen-

⁴³ “Eliminative Materialism and the Propositional Attitudes”, *Journal of Philosophy* 78 (1981): 67-90. La cita es de la p. 73.

sional... la rica variedad de ilusiones perceptuales... el milagro de la memoria... la naturaleza del proceso de aprendizaje mismo...

Churchland señala que hay fenómenos que se explican adecuadamente en el marco de la ciencia cognitiva pero que no son siquiera tocados por la ordinaria. Una suposición implícita adicional es que la ciencia cognitiva puede explicar todo lo que explica la psicología ordinaria. La pretensión es pues que la psicología ordinaria es reducible a la manera de Kemeny-Oppenheim a la ciencia cognitiva sistemática.

Churchland también es escéptico acerca de la reducibilidad conservadora nageliana de la psicología ordinaria a la ciencia cognitiva. Dice: "Pienso que no puede descartarse una reducción exitosa, pero la impotencia explicativa y el prolongado estancamiento de la PP inspiran poca fe en que sus categorías se encuentren claramente reflejadas en el marco de la neurociencia".⁴⁴ Aquí está diciendo que es improbable una reducción conservadora porque hay pocas razones para creer en la existencia de leyes puente que conecten los estados de la psicología ordinaria con los estados neurofisiológicos. Por tanto, la estructura de la argumentación de Churchland ejemplifica el patrón que discernimos con anterioridad: la psicología ordinaria debe ser eliminada porque es reducible a la neurociencia cognitiva al modo de Kemeny-Oppenheim sin que sea reducible a la manera conservadora.⁴⁵

En consecuencia, nos enfrentamos con la siguiente pregunta: dado que la neurociencia posee un rango explicativo más amplio que la psicología ordinaria, ¿por qué no pueden de todos modos coexistir, sin que la psicología ordinaria sea reducida nomológicamente a la neurociencia? ¿Por qué deberíamos descartar a la ordinaria y concluir que no hay, y nunca ha habido, cosas tales como creencias, anhelos, pesares y deseos?

El principio de exclusión explicativa proporciona una explicación simple de por qué ambas teorías, aun cuando no se excluyen lógicamente o empíricamente [*evidentially*], compiten entre sí y de por qué su convivencia pacífica es una ilusión. Dado que tanto la psicología ordinaria como la neurociencia pretenden proporcionar explicaciones para el mismo dominio de fenómenos y dado el fracaso de la reducción en cualquier dirección, las

⁴⁴ *Ibid.*, p. 75.

⁴⁵ Churchland también sugiere que las leyes de la psicología ordinaria han sido refutadas. Este es un punto controvertible y, aun cuando sea verdadero, no fuerza a la *eliminación* de la psicología ordinaria (al menos, la de la psicología intencional); todo lo que mostraría es que la psicología ordinaria necesita mejoramiento. Para argumentar en favor de una eliminación lisa y llana, algún principio tal como el de exclusión explicativa parece esencial, y si tenemos tal principio, no necesitamos la premisa que afirma la falsedad de la psicología ordinaria.

pretendidas explicaciones deben ser consideradas independientes. En consecuencia, de acuerdo con el principio de exclusión explicativa, una de ellas debe desaparecer.⁴⁶

Pienso que consideraciones similares pueden dar cuenta de un aspecto de otro modo enigmático de la famosa teoría de los “paradigmas” científicos de Thomas Kuhn.⁴⁷ De acuerdo con Kuhn, los sucesivos paradigmas que conciernen al mismo rango de fenómenos son “incommensurables” entre sí. Hacen uso de diferentes conceptos, diferentes metodologías, diferentes criterios para la generación de problemas y la evaluación de las soluciones propuestas. Como bien se sabe, Kuhn escribe en varios lugares que los paradigmas diferentes ni siquiera comparten los mismos problemas y quizá no pueden hacerlo; tampoco comparten en sentido estricto los datos. Pero aquí estoy suponiendo que Kuhn debe admitir un sentido en el que diferentes paradigmas sucesivos pueden compartir, y lo hacen, un dominio de cuestiones que se superponen. De otro modo, gran parte de su teoría de los paradigmas tiene poco sentido.

Ahora se podría plantear la siguiente pregunta acerca de los paradigmas: si son, como dice Kuhn, mutuamente incommensurables y por tanto no pueden siquiera contradecirse entre sí, ¿por qué no podemos aceptarlos a todos? ¿Por qué tenemos que descartar el paradigma antiguo cuando elaboramos uno nuevo? Después de todo, según Kuhn, ningún paradigma es literalmente refutado y todo paradigma es útil para ciertos propósitos explicativos y predictivos, realizando sus contribuciones científicas únicas. Por tanto, ¿por qué no acumular paradigmas? Podemos hacerlo sin temor a la incoherencia o la inconsistencia lógicas, pues los paradigmas son mutuamente incommensurables y por tanto no pueden contradecirse entre sí. Tendríamos una “teoría acumulativa” del progreso científico a través de la acumulación de paradigmas en lugar de la usual teoría acumulativa del conocimiento científico que Kuhn rechaza.

De nuevo, falta aún una respuesta desde el principio de exclusión explicativa: Kuhn da por sentada la incompatibilidad de paradigmas sucesivos acerca de los mismos fenómenos porque tácitamente acepta el princi-

⁴⁶ En mi opinión, la manera correcta de salvar la psicología ordinaria es dejar de pensar que ella juega el mismo partido que se supone que juega la “ciencia cognitiva”; es decir, dejar de pensar en ella como una “teoría” cuya razón de ser fundamental es generar predicciones y explicaciones causales basadas en leyes. Haremos mejor si nos centramos en su rol normativo en la evaluación de las acciones y la formación de las intenciones y decisiones. Si la psicología ordinaria compite con la ciencia cognitiva en el juego predictivo, no puede ganar, y lo máximo que puede esperar es su absorción de manera reductiva por rival más sistemático (y mejor fundado).

⁴⁷ *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

pio de exclusión explicativa, y lo seguimos porque nosotros tampoco lo cuestionamos. Considero que para Kuhn cada paradigma pretende proporcionar explicaciones completas e independientes de los datos en su dominio; completas e independientes en relación con otros paradigmas competitivos. Esto se sigue del principio de exclusión explicativa: no más de un paradigma para un dominio único.⁴⁸

BROWN UNIVERSITY

BIBLIOGRAFIA

Peter Achinstein, *The Nature of Explanation* (New York y Oxford: Oxford University Press, 1983).

Sylvain Bromberger, "An Approach to Explanation", en R. J. Butler (comp.), *Analytic Philosophy*, 2nd series (Oxford: Basil Blackwell, 1965).

Paul Churchland, *Scientific Realism and the Plasticity of Mind* (Cambridge: Cambridge University Press, 1979).

Paul Churchland, "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy* 78 (1981): 67-90.

Robert C. Cummins, *The Nature of Psychological Explanation* (Cambridge: The MIT Press, 1983).

Donald Davidson, "Mental Events", en Davidson, *Essays on Actions and Events* (New York: Oxford University Press, 1980).

Daniel C. Dennett, "Mechanism and Responsibility", en Dennett, *Brainstorms* (Montgomery, Vt.: Bradford Books, 1981).

Fred Dretske, "Referring to Events", en Peter A. French *et al.* (comps.), *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: University of Minnesota Press, 1977).

Michael Friedman, "Explanation and Scientific Understanding", *Journal of Philosophy* 71 (1974): 5-19.

Alvin I. Goldman, "The Compatibility of Mechanism and Purpose", *Philosophical Review* 78 (1969): 468-82.

Alvin I. Goldman, *A Theory of Human Action* (Englewood Cliffs, N. J.: Prentice-Hall, 1970).

Carl G. Hempel, *Aspects of Scientific Explanation* (New York: The Free Press, 1965).

Terence Horgan y James Woodward, "Folk Psychology Is Here To Stay", *Philosophical Review* 94 (1985): 197-226.

⁴⁸ He recibido comentarios y sugerencias útiles de David Benfield, John Biro, Brian McLaughlin, Joseph Mendola y Michael Resnik.

John Kemeny y Paul Oppenheim, "On Reduction", *Philosophical Studies* 7 (1956): 6-19.

Jaegwon Kim, "Causation, Nomic Subsumption, and the Concept of Event", *Journal of Philosophy* 70 (1973): 217-36.

Jaegwon Kim, "Events as Property Exemplifications", en Myles Brand y Douglas Walton (comps.), *Action Theory* (Dordrecht: Reidel, 1976).

Jaegwon Kim, "Causality, Identity, and Supervenience in the Mind-Body Problem", *Midwest Studies in Philosophy* 4 (1979): 31-49.

Jaegwon Kim, "Causes as Explanations: A Critique", *Theory and Decision* 13 (1981): 293-309.

Jaegwon Kim, "Concepts of Supervenience", *Philosophy and Phenomenological Research* 65 (1984): 257-70.

Jaegwon Kim, "Epiphenomenal and Supervenient Causation", *Midwest Studies in Philosophy* 9 (1984): 257-70.

Jaegwon Kim, "Self-Understanding and Rationalizing Explanations", *Philosophia Naturalis* 21 (1984): 309-20.

Jaegwon Kim, "Psychophysical Laws", en Ernest LePore y Brian McLaughlin (comps.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson* (Oxford: Basil Blackwell, 1985).

Jaegwon Kim, "Explanatory Realism, Causal Realism, and Explanatory Exclusion", por aparecer en *Midwest Studies in Philosophy*.

Philip Kitcher, "Explanation, Conjunction, and Unification", *Journal of Philosophy* 73 (1976): 207-12.

Thomas Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

David Lewis, "Causation", *Journal of Philosophy* 70 (1973): 556-67.

David Lewis, "Causal Explanation" en *Philosophical Papers II* (New York: Oxford University Press, 1986).

Alasdair C. MacIntyre, "Determinism", *Mind* 66 (1957): 28-41.

Norman Malcolm, "The Conceivability of Mechanism", *Philosophical Review* 77 (1968): 45-72. Reimpreso en Gary Watson (comp.), *Free Will* (Oxford: Oxford University Press, 1982).

Ernest Nagel, *The Structure of Science* (New York: Harcourt, 1961).

Peter Railton, "Probability, Explanation, and Information", *Synthese* 48 (1981): 233-256.

William L. Rowe, "Neurophysiological Laws and Purposive Principles", *Philosophical Review* 81 (1971): 502-508.

Michael Scriven, "Causation as Explanation", *Nous* 9 (1975): 3-16.

Ernest Sosa, "Mind-Body Interaction and Supervenient Causation", *Midwest Studies in Philosophy* 9 (1984): 271-81.

Stephen P. Stich, *From Folk Psychology to Cognitive Science* (Cambridge: The MIT Press, 1983).

G. H. von Wright, *Explanation and Understanding* (Ithaca: Cornell University Press, 1971).

ABSTRACT

A discussion and defense of "the principle of explanatory exclusion" —no event can be given more than one complete and independent explanation— this paper reopens a debate between N. Malcolm and A. Goldman: Can the same bit of behavior be given both a "purposive" (or "belief-desire" or "folk" psychological) and a "mechanistic" (or "cognitive" or "neuro-" scientific) explanation? Yes, the author holds, but at least one of the explanations will be dependent (reductively or by supervenience) on the other or incomplete in itself. Wanting an explanation for an event is an epistemic predicament, but having too many explanations is inherently unstable, generating a search for the relationship between the explanations. The *metaphysical* principle of explanatory exclusion is distinguished from the corresponding *epistemological* exclusion principle and the relationship between explanatory exclusion and explanatory realism considered. Finally, the article examines implications of explanatory exclusion for eliminativism and for Kuhn's theory of scientific "paradigms".