

Universidad Nacional de Mar del Plata

Facultad de Humanidades

Departamento de Filosofía

Maestría en Epistemología y Metodología de la Ciencia

Tesis de Maestría:

**Problemas epistemológicos y ontológicos del
principio de exclusión explicativa**

Autor: Lic. Gustavo Fernández Acevedo

Mat. N° 141/94

Director: Prof. Alicia Gianella

Abril de 1998

INDICE

INTRODUCCIÓN	4
I. EL PROBLEMA DE LA EXCLUSIÓN: ANTECEDENTES Y ALCANCES	9
1. LA POSIBILIDAD DEL MECANICISMO: EL PLANTEO DE MALCOLM	9
2. LA COMPATIBILIDAD ENTRE MECANISMOS Y PROPÓSITOS: LA RÉPLICA DE GOLDMAN ...	13
3. EL PRINCIPIO DE EXCLUSIÓN PARA LAS EXPLICACIONES CAUSALES	16
4. CAUSACIÓN MENTAL: EL NAUFRAGIO DEL MATERIALISMO NO REDUCTIVO	20
5. UN ARGUMENTO GENERAL EN FAVOR DE LA EXCLUSIÓN EXPLICATIVA	28
6. PLANTEOS ALTERNATIVOS Y VINCULADOS DEL PROBLEMA	30
II. FUNDAMENTOS ONTOLÓGICOS DE LA EXCLUSIÓN	39
1. PLURALISMO EXPLICATIVO, CAUSALISMO EXPLICATIVO Y PRINCIPIO DE EXCLUSIÓN: UNA PARADOJA	39
2. ¿PUEDE EVITARSE LA EXCLUSIÓN?	42
III. SUPUESTOS E IMPLICACIONES EPISTEMOLÓGICAS DEL PRINCIPIO DE EXCLUSIÓN: UNA CRÍTICA	52
1. CONSECUENCIAS EPISTEMOLÓGICAS EN EL PLANTEO DE MALCOLM	53
2. LOS ALCANCES EPISTEMOLÓGICOS DEL PRINCIPIO DE EXCLUSIÓN	55
3. ¿ES EL PRINCIPIO DE EXCLUSIÓN NEUTRAL CON RESPECTO A LAS TEORÍAS DE LA EXPLICACIÓN?	58
4. EL PRINCIPIO DE EXCLUSIÓN Y EL RECHAZO DEL REALISMO EXPLICATIVO	61
5. EL PRINCIPIO DE EXCLUSIÓN Y LA ELECCIÓN DE TEORÍAS	68
III. EL CASO DE LA <i>FOLK PSYCHOLOGY</i>	78
1. EL PRINCIPIO DE EXCLUSIÓN COMO FUNDAMENTO DE LA ELIMINACIÓN	79
2. CHURCHLAND Y LOS ARGUMENTOS EN FAVOR DEL ELIMINATIVISMO	81
3. UNA DEFENSA EPISTEMOLÓGICA DE LA PSC	87

4. UNA DEFENSA FUNCIONALISTA DE LA PSC	89
5. ¿QUÉ LUGAR QUEDA PARA EL PRINCIPIO DE EXCLUSIÓN?	92
CONCLUSIONES	95
REFERENCIAS BIBLIOGRÁFICAS	98

INTRODUCCIÓN

Mi propósito en el presente trabajo es analizar ciertos problemas recientes en filosofía de la mente y filosofía de la ciencia; en particular, me interesa examinar los fundamentos y consecuencias del denominado "problema de la exclusión causal-explicativa". Este problema, de creciente importancia especialmente para la primera de las áreas mencionadas, ha sido estudiado por diversos autores, pero ha sido en especial Jaegwon Kim quien ha analizado y sistematizado sus consecuencias epistemológicas. Si bien fue planteado explícitamente en las últimas décadas, hunde sus raíces en los debates relativos al problema mente-cerebro del último medio siglo, que reseñaré brevemente.

A la caída en desgracia que sufrió en nuestro siglo el dualismo de sustancias,¹ le siguieron diversos intentos de eliminar a lo mental, o, en el mejor de los casos, reducirlo de forma tal de ocupar un lugar marginal dentro de la ontología. De esta forma se observó el surgimiento del conductismo lógico o filosófico, la teoría de la identidad mente-cerebro y el materialismo eliminativo. Por diversas razones (cfr. Bechtel, 1988) el conductismo filosófico no obtuvo demasiada aprobación, y el materialismo eliminativo y la teoría de la identidad (o más precisamente teoría de la identidad como tipo) se constituyeron en las alternativas preferidas para estudiar filosóficamente los fenómenos mentales. Si bien el materialismo eliminativo gozó de alguna aceptación por parte de filósofos de prestigio (Rorty y Feyerabend, entre otros), fue la teoría de la identidad, en la forma que le dieron autores como Herbert Feigl y J.J.C. Smart, la que durante las décadas del '50 y '60 recogió el interés de la

mayoría de los filósofos ocupados en el tema. Su reinado duró hasta aproximadamente fines de la década del '60, cuando una serie de cuestionamientos pusieron en tela de juicio su adecuación como teoría acerca de lo mental. En especial, una serie de influyentes artículos atacaron los fundamentos mismos del programa, aparentemente revelándolo como inviable.² Esta presunta insatisfactoriedad de las posiciones dominantes llevó al desarrollo de concepciones que, dentro de un marco materialista y naturalista, intentaron preservar un lugar diferenciado para los fenómenos mentales, rechazando la posibilidad de una reducción o eliminación. Este programa general, que ha sido denominado "materialismo no reductivo" ha sido junto con el funcionalismo el marco filosófico dominante para el estudio de lo mental en las últimas décadas.

Pero a pesar de su aparente éxito en conciliar una ontología materialista con una perspectiva dualista de las propiedades mentales, no ha carecido de problemas internos. Uno de los primeros filósofos en plantear la gravedad de los problemas que su compromiso simultáneo con el fisicalismo y con la existencia de propiedades mentales no reductibles acarrea ha sido justamente Kim. El planteo sobre la exclusión de Kim comienza con un análisis de la imposibilidad de que las propiedades mentales tengan poderes causales sobre la conducta, de lo cual se deriva su irrealdad e irrelevancia explicativa. Esta conclusión básica le servirá luego como "caso testigo" para la formulación de un principio de exclusión explicativa mas general. En mi evaluación de la propuesta de Kim no me ocuparé en examinar en profundidad el argumento en favor de la ineficacia causal de las propiedades mentales, sino que

¹ Aunque pueden encontrarse intentos contemporáneos de revitalizar este enfoque. Por citar el mas conocido: Popper y Eccles, (1977).

² Kim (1989b) y Rabossi (1995) indican al clásico artículo de Putnam "The nature of mental states" (1967) como responsable de la decadencia de la teoría de la identidad. También son mencionados los célebres argumentos de Davidson en favor del anomalismo de lo mental en su artículo "Mental events" (1970). (Al respecto, *cf.* también Kim (1992)).

lo tomaré como marco para el análisis y generador de la formulación mas general del principio de exclusión.

El problema de la exclusión no sólo constituye una cuestión de importancia teórica para la filosofía de la mente y la epistemología; existe una aceptación creciente de la tesis de que la investigación filosófica en esta área tiene importantes repercusiones para el desarrollo de la psicología en general, y para la psicología cognitiva en particular. Coincido con Bechtel cuando sostiene que

existe a menudo una estrecha interacción entre las afirmaciones filosóficas y los esfuerzos de la investigación empírica de tal manera que aquellos que toman parte en una investigación empírica suponen frecuentemente, de manera conciente o inconciente, una postura filosófica particular (1988, p. 18).

Si esta afirmación es cierta, los desarrollos en los programas filosóficos hacen sentir sus efectos en los programas de investigación científica. Según algunos autores, en el caso de la filosofía de la mente y su vínculo con la psicología o la ciencia cognitiva, esta relación no es unidireccional: algunos problemas filosóficos pueden ser resueltos por la investigación científica. Pero aún cuando no se suscriba esta última tesis, queda en pie el hecho de que existe una interacción efectiva entre filósofos de la mente y científicos cognitivos, y es importante esclarecer los problemas básicos que se plantean a partir de esta interacción. Un buen ejemplo de la influencia de la reflexión filosófica en el área de la investigación cognitiva lo constituye el conocido argumento de Putnam en contra de la teoría de la identidad de tipo mente-cerebro, que forma parte del fundamento filosófico de los programas de investigación no reduccionistas de la mente.

En mi análisis de los fundamentos y argumentos en favor del denominado "principio de exclusión causal-

explicativa", corolario del problema homónimo, propondré y defenderé las siguientes tesis:

- 1.El principio de exclusión constituye una restricción general plausible para las explicaciones, si bien en el análisis general de Kim constituye una extensión injustificada de su formulación causal.
- 2.La versión epistemológica del principio de exclusión explicativa resulta aceptable de manera irrestricta sólo si se acepta además la versión metafísica.
- 3.Si bien Kim se refiere al principio como una "restricción general plausible", en ocasiones parece considerarlo una restricción necesaria para todo caso posible y para cualquier teoría de la explicación. En oposición a esto, sostendré que algunas teorías de la explicación pueden, en ciertos casos, admitir una violación del principio.
- 4.Los alcances epistemológicos que Kim asigna al principio en relación a la elección de teorías que compiten en un mismo dominio no están justificados desde un punto de vista teórico.
- 5.Los análisis *standard* del caso privilegiado que Kim examina (el de la psicología *folk*) sugieren, en concordancia con la tesis anterior, que debe utilizarse criterios mas amplios que el poder explicativo para la evaluación de teorías en competencia.

Estas tesis podrían sintetizarse en la afirmación de que, si bien el principio de exclusión explicativa constituye una restricción general plausible para las explicaciones, en el análisis de Kim se postula por una extensión insuficientemente justificada de afirmaciones plausibles de menor alcance, y que sus implicaciones y alcances epistemológicos son inferiores que las que pretende.

Para la defensa de estas tesis estructuraré mi trabajo en cuatro secciones. En el capítulo I expondré el planteo acerca del problema de la exclusión describiendo las

formulaciones anteriores del mismo, algunos planteos contemporáneos vinculados, y deteniéndome en particular en la propuesta de Kim. En el capítulo II comentaré una supuesta paradoja generada por el planteo de Kim, y discutiré la solución propuesta a la misma, consistente en rechazar cualquier tipo de principio de exclusión explicativa. En el capítulo III expondré una crítica detallada al planteo de Kim, señalando lo que considero son insuficiencias argumentativas y pretensiones infundadas acerca de los alcances del principio de exclusión; en este capítulo y en el anterior se concentra el núcleo de la argumentación. En el capítulo IV analizaré los argumentos en favor y en contra de la eliminación del marco intencional provisto por la denominada "psicología *folk*", mostrando que en los mismos no se hace uso (o se hace un uso muy limitado) de un principio de exclusión explicativa como el que Kim propone. Por último, en la Conclusión examinaré las tesis y los argumentos de Kim que no se han visto afectados por la crítica expuesta, y señalaré posibles líneas de continuación de la investigación.

I. EL PROBLEMA DE LA EXCLUSIÓN: ANTECEDENTES Y ALCANCES

Mi objetivo en el presente capítulo es ofrecer una exposición detallada del problema de la exclusión, comenzando con los planteos iniciales y deteniéndome luego en especial en la reformulación y ampliación que Kim hizo del mismo. No es mi propósito detenerme en consideraciones críticas hacia las posiciones expuestas, si bien ocasionalmente realizaré algunos comentarios sobre aspectos que crea que merecen algún análisis adicional. Tampoco realizaré un análisis exhaustivo de todos los enfoques contemporáneos sobre el tema, sino que me limitaré a describir en profundidad el análisis de Kim, que es el autor que posiblemente ha teorizado en forma mas amplia y argumentado en favor de un principio general de exclusión explicativa.

1. La posibilidad del mecanicismo: el planteo de Malcolm

Quizás el primer planteo explícito del problema de la exclusión (aunque sin denominarlo de esta forma) pueda situarse con cierta precisión en 1968,³ 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" (1968, 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

³ Kim cita como antecedente de una posición similar a la de Malcolm a MacIntyre (1957).

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 comprensiva, 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 hacen. Esta última 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 por medio de una escalera. Es el caso que el viento ha volado 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.⁴

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.⁵ Una segunda consecuencia de 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 falsados.

Un vez que ha especificado las diferencias entre los *explanans* de ambos tipos de explicaciones, Malcolm dirige su atención directamente a la posibilidad de que las mismas no sean realmente rivales. Esto es lo que han sostenido

⁴ 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.

⁵ Kim señala insuficiencias del concepto de dependencia que maneja Malcolm. En particular, señala que "la noción de 'dependencia' utilizada es demasiado estrecha y en el mejor de los casos parece caracterizar 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).

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.⁶ 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 la misma 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 con 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. Y si se pretendiera estipular que la localización de una determinada intención es la misma localización de sus procesos neurales correlativos, la identidad que resultaría de esta correlación ya no sería contingente.

⁶ 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. Volveré luego sobre esta cuestión.

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.⁷ Ante la existencia de una teoría tal, no quedaría espacio para las explicaciones propositivas: ambos tipos de explicación, como vimos, no son compatibles.

2. La compatibilidad entre mecanismos y propósitos: la réplica de Goldman

En un artículo en respuesta al de Malcolm, Alvin Goldman (1969) sostuvo la compatibilidad entre explicaciones propositivas y neurofisiológicas. El objetivo de Goldman era 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" (1969, p. 470. Las cursivas son de Goldman).

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, \dots, C_h son conjuntamente suficientes para la ocurrencia del suceso E , entonces ningún suceso salvo C_1, \dots, 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 en la misma es una causa del eslabón subsiguiente, y todos ellos pueden ser considerados una causa del eslabón final de la cadena. Dado que la exclusión

⁷ No obstante, señala dos aspectos en los cuales el mecanicismo *no es concebible*. *Infra*, p. 53.

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).⁸

Goldman considera que tanto del principio (I) como del (II) puede probarse que son inaceptables. Recurre 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 hubiera 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

⁸ 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.

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 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 como suficientes el uno para el otro.

3. El principio de exclusión para las explicaciones causales

En uno de los primeros artículos en los que se ocupó en profundidad del problema (1989a),⁹ 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.

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 (cosa 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*.

⁹ Si bien hay referencias anteriores a la exclusión al menos en su (1979) y (1987).

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.¹⁰

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 con contenido (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.

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

¹⁰ 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 a las misma 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.

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.¹¹

De acuerdo con Kim, el principio de exclusión explicativa se plantea de manera mas 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

¹¹ Coincido 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. Pienso que 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, veo difícil evitar la consecuencia de que toda la conducta humana está sobredeterminada. Goldman podría evitar esta consecuencia 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, ¿que manera tenemos 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. Creo que la solución de los equivalentes nómicos simultáneos -pienso que Kim coincidiría- debe tender a resolverse en un caso de sobredeterminación, un caso de dependencia asimétrica, o un caso de identidad.

de la causación psicofísica esta posibilidad toma la forma de la teoría de la identidad mente-cuerpo. Este caso, sostiene Kim, proporciona la manera mas simple y posiblemente mas 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 "sobreveniente" 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 la explicación 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 mas 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 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. Las cursivas son de Kim). Si esto es correcto, concluye, no es éste un caso en

el cual sean posibles dos explicaciones completas e independientes del mismo acontecimiento.¹²

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* se observa que pueden encontrarse (y, mas 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.

4. Causación mental: el naufragio del materialismo no reductivo

El análisis de Kim expuesto hasta el momento es bastante general, ya que las consideraciones relativas a las explicaciones causales de sucesos (y también las referentes a los equivalentes nómicos simultáneos y a los epifenómenos) no se aplican únicamente a los problemas

¹² 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 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 ampliamente el marco de esta tesis. Sin embargo, quisiera señalar que la observación de Kim con respecto a la "imagen engañosa e incompleta" de las explicaciones tomadas aisladamente para probar su presunta incompletud me parece notoriamente insuficiente e injustificada. Creo que en este caso sólo tenemos dos opciones: o aceptamos que hay casos de sobredeterminación causal y, por lo tanto, también sobredeterminación explicativa, o negamos que existan casos de sobredeterminación causal, con lo cual desaparecen los posibles casos de sobredeterminación explicativa. 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.

explicativos de las acciones humanas, sino que tienen un alcance mas vasto. En este apartado expondré la crítica de Kim al materialismo no reductivo, centrada en los problemas de la causación mental, (en particular en el problema de la "causación descendente"). La aplicación del principio de exclusión resultará crucial para el desarrollo y defensa de algunos aspectos del argumento. En el resto de este apartado me basaré principalmente en Kim (1993a).

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*. Por último, 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 *reducibles 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 reducibles 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 reducibles 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, afirma 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 reducibles 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 instancia 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 hablar¹³ (Kim, 1993a, p. 344).

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.

A pesar de que de este conjunto de supuestos parece desprenderse una imagen plausible de la mente y su lugar en una ontología general, los proponentes del materialismo no reductivo deben enfrentarse al problema de la exclusión, que surge a partir de los fundamentos mismos de su programa.

En la presentación de Kim, este problema admite dos versiones, una causal, y otra explicativa. La vertiente causal del problema de la exclusión para el debate sobre la

causación mental surge de los propios supuestos del programa materialista no reductivo. Debido a que sostiene la realidad de las propiedades mentales, el materialista no reductivo se compromete a sostener un rol causal activo de las mismas. Esto puede implicar, en opinión de Kim, tres casos posibles, siendo *M* una propiedad mental cuya instanciación causa: i) la instanciación de una propiedad *N*, que es una propiedad mental; ii) la instanciación de la propiedad *N*, que es una propiedad física; y iii) la instanciación de la propiedad *N*, que es una propiedad de nivel superior. Kim muestra que el problema surge de la siguiente forma:

Supóngase que *M* es causalmente eficaz con respecto a alguna propiedad mental *M**, y en particular que una instancia dada de *M* causa una instancia de *M**. Pero *M**, *qua* propiedad mental, es físicamente realizada; sea *P** su base de realización física. Ahora parece que tenemos dos respuestas independientes y distintas a la pregunta ¿por qué está presente ésta instancia de *M**? *Ex hypothesi*, está allí porque una instancia de *M* la causó; (...) Pero hay otra respuesta: está allí porque *P** realiza físicamente a *M** y *P** es instanciada en esta ocasión. Creo que estos dos relatos acerca de la presencia de *M** en esta ocasión crean una tensión, y esta tensión debe ser reconciliada (1993a, p. 351).

No es posible, prosigue Kim, que *M* y *P** constituyan conjuntamente una causa suficiente para la instanciación de *M**, dado que esto contradice la afirmación de que *P** instancia físicamente a *M**, y *P** por sí solo es suficiente para dicha instanciación. Y esta posibilidad también es inconsistente con la asunción anterior de que *M* por sí sola era suficiente para la instanciación de *M**. ¿Podríamos estar entonces en presencia de un caso de sobredeterminación? No, responde Kim, ya que si *M* y *P** fueran por sí mismas suficientes para causar *M**, entonces la presencia de *M* bastaría para causar *M** en ausencia de *P**, y esto sería contradictorio con la afirmación de que

¹³ Para Kim, este conjunto de supuestos configura una posición muy similar al emergentismo.

esa instancia de M^* está allí porque ha sido realizada por P^* . En base a esto, Kim considera que la única alternativa que un materialista no reductivo podría aceptar es la de suponer que la instancia de M causa M^* , causando la instanciación de P^* , su base de realización física. Esto es, por supuesto, causación descendente, de M a P^* , constituyendo un caso de causación de lo mental a lo físico. Kim sostiene que la esencia de su argumento queda encerrado en el principio siguiente:

[Principio de realización causal] Si una instancia dada de S ocurre al ser realizada por Q , entonces cualquier causa de esta instancia de S debe ser una causa de esta instancia de Q (y por supuesto cualquier causa de esta instancia de Q es una causa de esta instancia de S) (*ibid.*, p. 352).

No obstante, continúa Kim, tal vez podría evitarse el compromiso con la causación descendente explotando el hecho de que M , en tanto que propiedad mental, tiene su propia base de realización física P . Dado esto, podría pensarse que la causación de P^* por parte de M deviene en un caso de causación de lo físico a lo físico, esto es, M es físicamente realizado por P , y P causa P^* . La relación causal básica se da entonces entre dos propiedades físicas, P y P^* , y la causación de M^* por M está basada, en última instancia, en la relación causal existente entre sus bases de realización física. Kim considera que este es un punto de vista altamente atractivo, si bien considera que es inaceptable para un materialista no reductivista. Sería inaceptable porque reduce los poderes causales de M a los de su base de realización física P , y esto es contrario al *dictum* de Alexander y al supuesto de que M es una propiedad irreductible.

El materialista no reductivo se enfrenta entonces, según Kim, con la tensión derivada de combinar una "determinación ascendente" con una "causación descendente". Pretende que lo mental esté determinado por, y sea

dependiente de, lo físico, y a la vez que manifieste poderes causales nuevos, los cuales sólo pueden ser logrados por medio de afectar causalmente a procesos físico-biológicos. Pero dado que los poderes causales de *M* parecen depender de los poderes causales de *P*, habría motivos para considerar a *M* un epifenómeno. Si, a pesar de todo, se quisiera mantener la afirmación de que *M* tiene poderes causales distintos a los de *P*, se debería enfrentar a dos objeciones de peso. En primer lugar, dice Kim, el antiguo principio de simplicidad: si tenemos a *P* como causa de *P**, ¿por qué necesitamos a *M*? En segundo lugar, y mas grave aún, si se quisiera mantener a *M* como causa de *P**, el materialista no reductivo tendría que enfrentar el problema de la exclusión causal-explicativa. Resultaría absurdo, dice Kim, sostener que éste constituye un caso de sobredeterminación. En segundo lugar, *ex hypothesi*, resulta imposible que *P* y *M* constituyan, conjuntamente, una causa suficiente para la ocurrencia de *P**. Por último, dada la supuesta irreductibilidad de *M*, no puede decirse que sea idéntica a *P*, o parte de ella. El problema de la exclusión para las propiedades mentales se plantea entonces en los siguientes términos: dado que *P* es una causa suficiente de *P**, ¿cómo puede ser también *M* una causa suficiente de *P**? La afirmación de *M* como causa de *P** se debilitará especialmente si, como cabría esperar, la investigación neurobiológica muestra la existencia de una cadena causal continua, un mecanismo, que conecta *P* y *P**. Resulta claro, concluye Kim, que el problema de la exclusión no puede ser resuelto dentro del marco del materialismo no reductivo.

Como consecuencia de este planteo se advierte una conclusión indeseable para el materialismo no reductivo: lo mental es *causalmente ineficaz*. Esto, sumado a la definición de lo real aceptable para el materialista no reductivo (el *dictum* de Alexander) obliga a aceptar una conclusión inadmisibles: lo mental es necesariamente *irreal*,

lo cual contradice uno de sus presupuestos básicos.¹⁴ Esta ineficacia causal de lo mental conduce a la *irrelevancia explicativa causal* de lo mental (si lo mental es causalmente inerte, no puede tener lugar en explicaciones causales adecuadas).

El problema de la exclusión explicativa de las propiedades mentales surge entonces a partir de las consideraciones anteriores. La exclusión explicativa de estas propiedades como *explanans* apropiado se desprende del resultado precedente con respecto a su ineficacia causal. Si las propiedades mentales son causalmente ineficaces, no pueden constituir una competencia para explicaciones alternativas basadas en la eficacia causal de las propiedades físicas. Las explicaciones basadas en las propiedades mentales son, por lo tanto, dependientes de las explicaciones basadas en las propiedades físicas. Esta posición presupone, como se advertirá, una posición realista¹⁵ con respecto a la explicación científica.

Ahora bien, según Kim un principio general de exclusión explicativa (del cual la exclusión explicativa de las propiedades mentales constituye un caso particular) puede ser defendido sin necesidad de apelar a consideraciones causales, ni a la idea de que sólo son genuinas explicaciones las explicaciones causales.¹⁶ Mas aún, como veremos, Kim sostiene que ni siquiera hace falta

¹⁴ Sabatés (1995) ha propuesto un argumento para evitar la conclusión de que lo mental es irreal. Según él, no puede evitarse la conclusión de que lo mental es causalmente ineficaz y, mas aún, está causalmente aislado, con lo cual rechaza la alternativa epifenomenista. No obstante, sostiene que la realidad de las propiedades mentales puede conservarse por medio de rechazar el criterio del poder causal como criterio de realidad, y reemplazarlo por un criterio mas amplio. No me extenderé en el análisis de esta cuestión ya que, como él mismo advierte, esta propuesta proporciona una solución para el problema de la irrealidad, pero no para el problema de la causación mental. Por lo tanto, persiste el problema de la exclusión explicativo-causal para las propiedades mentales.

¹⁵ "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" (Kim, 1987, p. 226).

¹⁶ La aceptación por parte de Kim de que existen explicaciones que no son causales (lo que se ha dado en denominar "pluralismo explicativo") trae aparejado un presunto problema de consistencia interna en su defensa del principio de exclusión. *Infra*, p. 39 y ss.

adoptar una posición realista con respecto a la explicación científica para adoptar el principio de exclusión.

5. Un argumento general en favor de la exclusión explicativa

De acuerdo con Kim, cuando buscamos una explicación nos encontramos en un *predicamento epistémico*, estado del cual podemos salir por medio de una explicación exitosa. La posesión de dos o más explicaciones, por el contrario, nos volverá a colocar en una posición de predicamento epistémico, de la cual no saldremos hasta saber cómo se relacionan las explicaciones entre sí: 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" (1989a, p.254). La exclusión explicativa es entonces defendida por Kim a partir de argumentos tales como la *simplicidad* o la *unificación*¹⁷ que proporciona el contar con una única explicación de un hecho. Si la unidad y la simplicidad de la teoría constituyen nuestra meta cuando buscamos explicaciones, dice Kim, entonces la existencia de múltiples explicaciones de un mismo *explanandum* es contraproducente, a menos que se pueda determinar de una manera apropiada como los distintos *explanans* se relacionan entre sí. Si las premisas de una explicación dependen de las de otra (por ejemplo, en el sentido que Kim explicita para las explicaciones causales),¹⁸ es claro que las explicaciones no son independientes; y, en este caso, no estamos comprometidos con más suposiciones independientes acerca del mundo que lo que lo estamos aceptando una de las explicaciones. Así como la referencia a la situación de dependencia entre

¹⁷ Siguiendo una idea desarrollada por Michael Friedman y posteriormente por Philip Kitcher.

explicaciones causales suponía el realismo causal, esta posición general supone el realismo explicativo.

Si, en su versión causal, el principio de exclusión sostiene que no puede haber *mas de una causa* para un mismo hecho o suceso, en su versión explicativa, entonces, el principio de exclusión postula que no puede haber dos (o mas) 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" (*ibid.*, 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 aún 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 habérselas con dos (o mas) explicaciones independientes del mismo fenómeno. No es necesario que los respectivos *explanans* sean lógicamente contradictorios, y aún 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, las cursivas son del original).

¹⁸ *Supra*, pp. 18-20.

En una aplicación de las conclusiones precedentes, Kim menciona ciertas consecuencias epistemológicas con respecto al debate acerca de la psicología *folk* y su relación con una ciencia madura acerca de lo mental. Sostiene Kim que si las afirmaciones de autores como Paul Churchland son correctas, acerca de que una ciencia desarrollada acerca de lo mental (sea ésta ciencia cognitiva o neurociencia), puede explicar mas fenómenos que la psicología ordinaria, y, en los dominios que coinciden, explicar mejor, el principio de exclusión proporciona una razón por la cual no pueden coexistir pacíficamente. Las explicaciones proporcionadas por ambas no necesitan, como se mencionó mas arriba, ser contradictorias lógica o empíricamente. Pero dado que sus dominios explicativos se superponen, el principio de exclusión proporciona un argumento para sostener que una de las dos debe desaparecer. Retomaré este planteo en el capítulo IV.

6. Planteos alternativos y vinculados del problema

El planteo de Kim, si bien parece ser uno de los mas amplios y de mayor alcance relativos al problema de la exclusión, no es el único análisis sobre el mismo.

Si bien el enfoque es distinto, pueden encontrarse un análisis vinculado en Block (1990). Para Block, el problema se plantea en términos de si el funcionalismo puede evitar el epifenomenismo, entendiendo esta doctrina como la que sostiene que "lo que pensamos o queremos no tiene relevancia causal para lo que hacemos" (p. 29). Block formula 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. Si bien el problema planteado por Block guarda parentesco con el postulado por Kim, sus alcances no son idénticos (y tampoco tan amplios), por lo que no lo abordaremos en esta oportunidad.

Un planteo vinculado con el problema de la exclusión, pero tampoco idéntico a él, es el analizado por Kim en su (1991). En este artículo, Kim analiza la propuesta de Dretske acerca de cómo las razones explican. Dado que Dretske asume, con Davidson, que las explicaciones racionalizadoras son una clase de explicaciones causales, el problema se formula en términos de cómo los estados intencionales pueden ser causas. Esto es, el problema de la causación intencional o racional. Este es, dice Kim, un caso especial del problema más general de la causación psicofísica (de cómo los estados mentales pueden entrar en relaciones causales con estados físicos), y el cual es, a su vez, parte del problema más amplio de la causación mental (de cómo los estados mentales pueden entrar en cualquier tipo de relaciones causales). El enfoque de Dretske, según Kim, está destinado a responder a uno de los tres problemas fundamentales de la causación mental. Estos son el del anomalismo de lo mental, el sintacticalismo y el problema de la exclusión causal-explicativa. Estos tres problemas están, para Kim, estrechamente relacionados, por lo que un enfoque completo de la causación mental debería proveer una solución para los tres. Por "sintacticalismo" se entiende la doctrina que sostiene que sólo las propiedades sintácticas de estados internos, y no sus propiedades semánticas (o "contenido") son psicológicamente relevantes. Dado que además se acepta que el contenido, o significado, constituye una propiedad extrínseca que

involucra hechos referentes a las condiciones históricas y ecológicas del organismo, el problema de la causación mental generado por el sintacticalismo sería entonces responder a la pregunta ¿cómo pueden las propiedades extrínsecas ser causalmente eficaces en la producción de la conducta? Según Kim, Drestke (1990) adopta una solución del tipo "*explanandum* dual", para tratar con el problema de la relevancia causal del contenido; no obstante, afirma que esta posición, además de presentar varios problemas internos, no resuelve el problema de la exclusión causal-explicativa.¹⁹

Un análisis mas estrechamente relacionado al problema de la exclusión lo constituye el planteo de Jackson y Pettit (1990), quienes tratan de probar que hay una manera en la cual las propiedades pueden ser *causalmente relevantes*, y por lo tanto participar en explicaciones causales, sin ser causalmente eficaces. Dado que el intento podría ser una alternativa de solución al problema de las propiedades mentales planteado en los términos de Kim, lo describiré con cierto detalle. No obstante, es necesario hacer una aclaración: si bien, en el caso de que la propuesta fuese correcta, podría constituir un principio de solución a uno de los casos mas importantes abarcados por el principio de exclusión, no implicaría un rechazo total al mismo. Como vimos, los alcances del principio de exclusión son, para Kim, mucho mas amplios que el del problema restringido de las propiedades mentales.

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 en tales

¹⁹ Cfr. Kim, 1990, y 1991. En el volumen Macdonald (1995) Dretske expone de manera resumida su propuesta, junto con una réplica a Kim. La estrategia del *explanandum* dual tiene un precedente en von Wright (1970).

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 instanciación, 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. la F-instancia [*F-instance*] no ayuda a producir la G-instancia [*G-instance*], si G es eficaz, ayuda a producir e; F y G no son factores causales secuenciales; c. la F-instancia no se combina con la G-instancia, 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 descrito, es obvio que F no puede ser tomada como eficaz en el sentido en que lo es G. La relación entre la instanciación de F y la ocurrencia de e es secundaria a la relación entre la instanciación de G y la ocurrencia de e.

A las tres suposiciones descritas 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. 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 ciencia básica, presumiblemente la física. Esto se debería a que sólo las explicaciones en este último dominio harían referencia a propiedades causalmente eficaces.

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 es realizada y, en las circunstancias, que el suceso, bajo una cierta descripción, ocurre" (1990, p. 114).²⁰ Habría entonces al menos dos maneras distintas en las que una propiedad podría ser causalmente relevante: por medio de ser eficaz en la producción de un determinado efecto, o "programando" [*programming for*] la presencia de una propiedad eficaz. En este caso, la propiedad-instancia no participa en el proceso productivo que conduce al suceso, pero asegura que

²⁰ "The realization of the property ensures (...) that a crucial productive property is realized and, in the circumstances, that the event, under a certain description, occurs."

figure que la propiedad-instancia que es requerida para tal proceso. Los autores describen el rol de esta propiedad de manera metafórica, diciendo que su realización "programa" para la aparición de la propiedad productiva y, bajo una cierta descripción, para el suceso producido. Escriben al respecto

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).

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" [*program explanation*] puede tener una significación que permanezca aún 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.

En la descripción de Jackson y Pettit hay dos aspectos independientes para analizar, aunque íntimamente relacionados. El primero de ellos, referente a la medida en que una propiedad causalmente ineficaz puede ser causalmente relevante de forma tal que pueda servir de fundamento para una explicación. El segundo, relativo a la presunta información adicional que una explicación tal aportaría. Los analizaré sucesivamente.

Las consideraciones causales, obviamente, son las primeras que hay que abordar. *Prima facie* surge un problema que despoja al enfoque expuesto de su atractivo potencial. Hay una consideración en favor de la exclusión explicativa

causal para las propiedades mentales que no parece verse afectada por el argumento de Jackson y Pettit. Para afirmar la existencia de las explicaciones psicológicas como características de una "ciencia especial" queremos que tales explicaciones hagan referencia a propiedades *causalmente eficaces*, y no sólo causalmente relevantes. Y las "explicaciones por programa" sólo pueden ofrecernos referencia a propiedades causalmente relevantes en el sentido de "programar para" la presencia de propiedades eficaces, pero no eficaces en sí mismas. En el caso de las explicaciones psicológicas, según los autores, tendríamos explicaciones que "invocan contenidos actitudinales como propiedades causalmente relevantes" (*ibid.*, p. 115). Pero estos "contenidos actitudinales" no serían en sí mismos propiedades causalmente eficaces. Lo mismo ocurriría con las explicaciones de otras ciencias especiales. De hecho, Jackson y Pettit sugieren que tal vez muchas de las explicaciones que alguna vez podamos ofrecer serán explicaciones por programa, y que posiblemente sólo encontremos propiedades causalmente eficaces en la física. Este resultado obviamente no constituiría algo muy alentador para las pretensiones de un materialista no reductivo, por lo cual no podría ser considerado por éste como una solución satisfactoria; no defiende de manera suficientemente fuerte el rol productivo de las propiedades mentales.

Por otra parte, diría que las explicaciones "por programa" y "por proceso" de un mismo suceso no son independientes, en el sentido estipulado por Kim. En este caso tendríamos, de acuerdo con las posibilidades estudiadas por Kim en relación al principio de exclusión, dos explicaciones, pero parece ser el caso que una de ellas depende de la otra. La explicación por programa que refiere a propiedades mentales causalmente ineficaces resultaría en alguna medida dependiente de otra, basada en un nivel

inferior, tal vez físico, que se refiera a propiedades causalmente eficaces. Esto requiere algunas aclaraciones.

Para decir que una explicación es dependiente de otra, si se acepta una concepción realista de la explicación, es necesario que existe alguna clase especial de conexión nómica entre los correlatos objetivos de los respectivos *explanans*.²¹ Plantear que ambas explicaciones son dependientes en el modelo de Jackson y Pettit, entonces, requiere determinar la clase de relación existente entre F y G. Jackson y Pettit no son explícitos acerca de esta cuestión; un análisis de ese aspecto, por lo tanto, debería basarse en algunos comentarios no sistemáticos de los mismos y en el examen de los ejemplos planteados. No es mi propósito extenderme en profundidad en este análisis, así que sólo formularé observaciones esquemáticas. Sabemos, por definición (condición 3.1), que F y G son distintas, así que no se aplica la solución de la identidad. Pero esto deja varias posibilidades abiertas ¿es F causada por G? ¿o es F realizada físicamente por G? ¿o tal vez F sobreviene a G? Pero sea cual fuere la relación entre ellas, permanece en pie la idea de que pertenecen a niveles ontológicos distintos: F es una propiedad de orden superior, y G de orden inferior. Y dado que F es eficaz sólo si G es eficaz, se tendrá que reconocer entonces que una explicación que cuente a F en su *explanans* es una explicación sólo si existe también una explicación que incluya a G en su *explanans*; ambas explicaciones no son, por lo tanto, independientes

Con respecto a la presunta información adicional que las explicaciones por programa aportarían, si bien no puede descartarse que en algún caso pueda resultar relevante (cosa que, por otra parte, no ocurre en todos los casos, como los propios autores reconocen: decir que la fragilidad fue la causa de la rotura del vaso significa lo mismo que

²¹ Este es el sentido estipulado por Kim. *Supra*, p. 27n.

decir que una determinada estructura molecular lo fue), no resulta claro en general cual es la ganancia epistémica que supondría la información aportada por tales explicaciones. No me detendré en estas consideraciones; sólo quisiera decir que el análisis de este aspecto debería tomar en consideración las preguntas relativas a por qué debemos buscar explicaciones de los hechos que nos rodean, y a la clase de conocimiento del mundo que se espera que aporte una explicación. Sólo respondiendo previamente a estos interrogantes puede evaluarse la medida en que la información aportada por las explicaciones por programa puede resultar de interés.

En síntesis, el enfoque de Jackson y Pettit, si bien podría constituir un intento de interés en restituir a las propiedades causalmente ineficaces (entre ellas, las mentales) un lugar en la estructura causal del mundo, no alcanza a resolver el problema de la exclusión causal para las mismas. Y esto es así porque permanecen causalmente ineficaces: las propiedades físicas (o fisiológicas) son las causas de nuestra conducta, y no nuestras intenciones o creencias.

II. FUNDAMENTOS ONTOLÓGICOS DE LA EXCLUSIÓN

En el capítulo anterior se ha puesto de manifiesto que los argumentos de Kim en favor de la exclusión explicativa se basan principalmente en consideraciones ontológicas (si bien también defiende su aceptación sobre bases puramente epistemológicas). El análisis de la relación de causalidad juega un rol decisivo en esta argumentación. En este capítulo examinaré una crítica al planteo de Kim, que señala una inconsistencia en su propuesta a partir de la aceptación de ciertos supuestos que parecen correctos, y la presunta necesidad de renunciar al principio de exclusión explicativa para evitar tal inconsistencia. Sostendré que esta crítica fracasa en el logro de su objetivo; por el contrario, concluiré que el principio de exclusión explicativa constituye una "restricción general plausible" para las explicaciones, como consecuencia de un planteo ontológico de exclusión.

1. Pluralismo explicativo, causalismo explicativo y principio de exclusión: una paradoja

La idea de que existen explicaciones que no son causales se ha dado en denominar "pluralismo explicativo". Kim (1974, 1994b) admite la existencia de explicaciones basadas en relaciones no causales, tales como la relación de "dependencia mereológica" o "superveniencia mereológica". A partir de estos supuestos Sabatés (1996) ha señalado que la conjunción del causalismo explicativo y el pluralismo explicativo con el principio de exclusión explicativa conducen a la propuesta de Kim a una inconsistencia que puede ser resuelta mediante el rechazo del principio de exclusión, pero sin abandonar las intuiciones que lo motivan.

El núcleo del argumento de Sabatés puede resumirse en lo que sigue (dado que la posición de Kim con respecto al realismo explicativo y al principio de exclusión ya ha sido expuesta, mencionaré solamente los argumentos en favor del pluralismo explicativo). 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. La concepción metafísica de que el mundo está estructurado por diferentes relaciones de determinación o dependencia conduce a rechazar el causalismo explicativo. Un criterio más liberal, en su opinión, consiste en sostener el pluralismo explicativo (PE), de acuerdo con el cual existen, además de la relación causal, otras relaciones estructurantes en el mundo, tales como la dependencia o superveniencia mereológica,²² la identidad y la "dependencia de Cambridge". La existencia de las relaciones de superveniencia mereológica es suficiente, según Sabatés, para hacer al pluralismo explicativo más plausible que al causalismo explicativo estrecho.

El problema para Kim se presenta, de acuerdo con Sabatés, cuando a estos supuestos 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

²² La relación de superveniencia mereológica es la relación por la cual la instanciación de una propiedad en un objeto depende de las propiedades y las relaciones instanciadas entre sus partes (por ejemplo, la propiedad de ser transparente de un fluido depende de las propiedades y relaciones entre las moléculas del fluido).

existir mas que explicaciones causales para los sucesos del mundo, con lo cual el PE resultaría falso.

La solución que Sabatés propone a este problema consiste en rechazar el principio de exclusión explicativa, pero sin negar las ideas que subyacen al mismo. Sabatés señala con justicia, a mi modo de ver, que los argumentos de Kim en favor del principio de exclusión 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 (EC) sostendría que no pueden existir dos causas completas e independientes de un mismo suceso, mientras que el segundo (EEC) afirmaría que no puede haber dos explicaciones *causales* de un mismo suceso. Sabatés advierte que el principio general de exclusión explicativa de Kim consiste en una extensión injustificada a partir de EC, ya que, aún admitiendo que las consideraciones causales son automáticamente relevantes para las explicaciones, sus conclusiones se aplicarían solamente a las explicaciones causales. Dado que EC conduce solamente a EEC, concluye Sabatés, constituye un *non sequitur* el intento de utilizarlo para excluir algo mas 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 global de la realidad) no pueden construirse argumentos análogos para defender el principio general de exclusión explicativa. En particular, 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*. A esto podría responderse, dice Sabatés, 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. Esta réplica, dice Sabatés, implica lo que podría ser llamado "principio de exclusión de la dependencia" [*dependence-exclusion principle*] (PED): no puede haber dos determinantes suficientes para el mismo suceso. Pero si bien este principio es más fácil de obtener que el principio general de exclusión explicativa, sostiene Sabatés, es altamente implausible. Supóngase que el cambio en la concentración de iones de hidrógeno en una solución es un determinante suficiente para el cambio en el valor Ph de la solución. ¿Es el cambio en el Ph de la solución no causado? Si se tiene la convicción, prosigue, de que todo macro-suceso depende de sucesos microfísicos, entonces todo macro-suceso sería no causado. Por lo tanto, PED conduce a resultados inaceptables, por lo que no puede ser adoptado. Y sin PED, concluye, no existen bases para la exclusión al nivel explicativo.²³

2. ¿Puede evitarse la exclusión?

Creo que la crítica de Sabatés es acertada en cuanto a que las consideraciones ontológicas de Kim sobre la causalidad no son suficientes para sustentar adecuadamente a un principio general de exclusión explicativa. No obstante, considero que su análisis deja en las sombras el problema de la relación existente entre distintos tipos de relaciones estructurantes de la realidad (en particular,

²³ Sabatés advierte que los argumentos de Kim en favor del principio de exclusión basadas en consideraciones tales como la simplicidad y la unificación tampoco son suficientes para justificar su aceptación (*infra*, p. 63-64).

entre la causalidad y la superveniencia),²⁴ y que este problema debe ser abordado. Es posible que exista una manera de replantear los alcances del principio de exclusión de forma tal que respete el pluralismo explicativo dentro de una concepción realista de la explicación. Exploraré esta posibilidad mediante un ejemplo.

Supóngase que el agua contenida en un recipiente es expuesta a una fuente de calor hasta que entra en estado de ebullición. El estado de ebullición puede ser ontológicamente concebido como un estado macroscópico que está "basado" en un determinado estado microscópico (o en una serie de micro-sucesos) que constituye su base; en este caso, los movimientos violentos de las moléculas componentes del agua.

Sea e el estado que queremos explicar (la ebullición del agua); m el movimiento de las moléculas, y c la exposición del agua a la fuente de calor. ¿Cómo debería pensarse la relación existente entre m y c para producir e ? Podemos formular, con alguna modificación, la pregunta que Kim cree pertinente para determinar la existencia de genuinos casos de sobredeterminación causal. En este caso, podemos preguntar ¿habría acontecido e si, estando presente m , c no hubiera tenido lugar? Y también ¿habría acontecido e si, estando presente c , m no hubiera tenido lugar? Si la respuesta que se diera a ambas preguntas fuese negativa, no estaríamos en presencia de dos determinantes suficientes para la producción de un determinado estado. En ausencia del calentamiento no habría modificaciones en el movimiento de las moléculas, con lo cual el estado final (ebullición)

²⁴ Hablar de la noción de superveniencia "en general" es una decisión que puede acarrear varias dificultades. En primer lugar, no existe una única noción de superveniencia. Sólo en los trabajos de Kim relativos al tema se distinguen la superveniencia "débil", "fuerte" y "global" (1984), y cada una de ellas plantea una serie de problemas específicos. Además, existe una profusa bibliografía referente a esta relación, dado el interés que ha suscitado en campos muy disímiles de la reflexión filosófica. De esta forma, resulta muy difícil determinar en que medida el atribuir una serie de rasgos a la relación de superveniencia permite hablar con cierta seguridad de una noción de superveniencia que pueda ser compartida por los diferentes enfoques posibles.

no se produciría; por otra parte, en ausencia de un determinado estado de movimiento de las moléculas no se produciría la ebullición. En cualquier caso, en ausencia de uno de los dos determinantes no se observaría la presencia de un determinado estado macroscópico. Entonces, análogamente a un caso de causas parciales, estaríamos en presencia de un caso de *determinantes parciales*. Ambos determinantes, presuntamente, serían necesarios y conjuntamente suficientes para la ocurrencia del estado final observado.

Ahora supongamos que planteamos la pregunta ¿por qué está el agua en estado de ebullición? Si el análisis anterior fuese correcto, las explicaciones basadas en uno solo de los determinantes explicaciones *no constituirían explicaciones independientes y completas* del mismo *explanandum*, esto es, el estado de ebullición. Una de estas explicaciones aludiría al estado de las moléculas de agua, es decir, a un determinado tipo de movimientos caracterizados de una manera mas o menos definida. La segunda explicación haría referencia a la existencia de una fuente de calor a la que ha sido sometida el agua (como en el caso anterior, el *explanans* podrá tener el grado de precisión que se desee, en términos de tiempo a la que el agua fue sometida a la fuente de calor, las 'calorías' que esa fuente de calor proporcionaba, etc.). La primera explicación plantearía, como su correlato objetivo, a una relación no causal, sincrónica, entre un estado microscópico y un estado macroscópico; la segunda, una relación causal, temporal, entre la realización de un determinado suceso y la producción de otro como su efecto. Si se considerara que ninguno de los dos determinantes es suficiente para la producción del suceso, entonces una explicación completa debería aludir a los dos determinantes, que constituirían un conjunto suficiente para su producción. Si se aceptara este punto de vista, tal

vez el fundamento metafísico para la aceptación del principio de exclusión explicativa podría ser reformulado en la siguiente forma: "no puede existir mas de un conjunto suficiente de determinantes para un mismo suceso".

Pero creo que la respuesta de que ambos determinantes son necesarios para la producción del suceso (la ebullición) es errónea. Si decimos que la ebullición sobreviene a un determinado estado molecular, esto parece suficiente para la producción del suceso que se quiere explicar. A las preguntas formuladas anteriormente (¿habría acontecido e si, estando presente m , c no hubiera tenido lugar? ¿habría acontecido e si, estando presente c , m no hubiera tenido lugar?) se debería contestar afirmativamente a la primera y negativamente a la segunda. Podríamos entonces presentar una visión alternativa. En ella, el calentamiento hipotéticamente *causaría* una serie de sucesos microfísicos (el aumento de la velocidad de las moléculas), de los cuales, a su vez, *dependería* o *sobreveniría mereológicamente* un estado macroscópico (la ebullición). La relación entre estos determinantes podría ser descripta como una especie de "cadena", compuesta por eslabones causales y no causales, donde a un suceso producido como efecto de una determinada causa le sigue un suceso producido por una relación de dependencia mereológica con aquel. Bajo esta descripción, el último suceso suficiente para la existencia del estado macroscópico en la existencia de un determinado movimiento molecular, y no la existencia del calentamiento, que ocupa un lugar anterior en la cadena.²⁵ El problema con esta descripción radica claramente en que obliga a aceptar sucesos no causados: si la existencia de la "base" de dependencia es suficiente para la producción del suceso, entonces no hace falta ninguna causa para su producción. Si bien podría decirse

²⁵ Tal vez debería decirse "lógicamente anterior", ya que no hay intervalo temporal entre la producción del microsucedo (los movimientos moleculares) y la producción del macrosucedo (la ebullición).

que *c*, en el ejemplo, es *causalmente relevante* en la producción de *e*, no sería causa de su producción.

Si ahora preguntáramos si una explicación de la ebullición en términos del movimiento molecular es completa e independiente, tendríamos que responder afirmativamente. En caso contrario, cualquier explicación causal que no aludiera a eslabones anteriores en la cadena causal que conduce al efecto debería ser considerada incompleta, lo cual es altamente implausible. Si se acepta este planteo, hay que concluir que una explicación tal, dada la exclusión de la causación del suceso, *excluye* a una explicación en términos causales. Si, por el contrario, se considera que una explicación tal debe incluir un eslabón adicional (además del último) en la cadena, porque en caso contrario la explicación no sería completa, se abre la posibilidad de que una explicación que incluya referencias a la causa y a la superveniencia sea completa e independiente en sí misma.

Si se quiere preservar la idea de que todo suceso (micro o macro) tiene una causa, al parecer la descripción anterior no es compatible con este propósito. Si estas descripciones son todas las que hay, y queremos mantener algún principio de exclusión, tenemos dos opciones: o renunciamos a la idea de que todos los macro-sucesos tienen causas, o renunciamos a la idea de que las explicaciones por dependencia mereológica son completas. Coincido con Sabatés en que sin algún principio de exclusión de determinantes no hay bases para la exclusión al nivel explicativo. Entonces, hay que decidir si es preferible renunciar a la exclusión explicativa sobre la base de que no puede existir ningún principio metafísico de exclusión, o renunciar a la idea de que las explicaciones por dependencia mereológica son en sí mismas completas, o renunciar a la idea de que todo suceso tiene una causa.

Pero tal vez el problema podría plantearse en términos mas simples, tomando como elementos sólo a la relación de

supervenencia y a la relación causal, sin hacer referencia a ningún principio de exclusión. Supongamos que el pluralista explicativo se limita a decir que la única relación que puede servir de base a una explicación, además de la relación causal, es la de supervenencia mereológica. Ahora bien, un pluralista no se conforma con decir que hay explicaciones por dependencia mereológica. Quiere también afirmar que estas explicaciones *son en sí mismas independientes*. Si una explicación es completa e independiente, esto quiere decir que el determinante que produce el suceso a explicar (la relación de dependencia mereológica, en este caso) es *suficiente* para que el mismo aparezca. Pero, si no se quiere rechazar la afirmación de que todo suceso tiene una causa, entonces habría casos de sucesos que, además de estar causados, tienen un determinante en la forma de una relación de dependencia mereológica -lo cual sería una sobredeterminación no causal-.

Los ejemplos que desarrolla Sabatés no contribuyen, me parece, a aclarar esta cuestión. Transcribo uno de ellos, el cual, aunque bastante extenso, expresa estas dificultades

Supóngase que un suceso *b* explica un suceso *e* vía dependencia o supervenencia mereológica, y que un suceso *c* causa *b* y por lo tanto explica causalmente *b*. O menos abstractamente, supóngase que el cambio en el valor P_h en una solución (*e*) es explicado por un cambio en la concentración de iones de hidrógeno en la solución (*b*), y que ese cambio es causado (y causalmente explicado) por el hecho de que una sustancia particular ha sido añadida a la solución (*c*). ¿Es posible afirmar que la explicación mereológica depende de una causal? No puede ser sobre la explicación causal entre *c* y *b* porque, como ya hemos visto, en este sentido espurio toda explicación causal también dependería de explicaciones que refirieran a eslabones causales previos. La única salida es decir que *c* también causa *e* y argumentar que, dado que *c* explica causalmente tanto *e* como *b*, la explicación de *e* en términos de *b* es en algún sentido dependiente de un par de explicaciones causales. *Pero esto no puede ser. Efectos de la misma causa no están siempre (de hecho sólo ocasionalmente) explicativamente relacionados.* (...) Lo que

hace a la relación entre *b* y *e* (entre el cambio en el nivel de concentración de iones de hidrógeno en la solución y el cambio en el valor Ph) explicativa tiene que ser algo que no sea el caso en otras instancias de eventos que comparten la misma causa. Y este algo solamente puede ser la relación de dependencia mereológica entre *b* y *e*. Entonces, si lo que hace a la relación entre *b* y *e* explicativa es esta relación de dependencia mereológica, no podemos decir en ningún sentido sustancial que tal relación explicativa depende de una explicación causal (o de un par de explicaciones causales) (1996, p. 101. Las cursivas son mías).

Sabatés afirma claramente que no puede ser que la explicación de *e* en términos de *b* sea dependiente de un par de explicaciones causales, *pero no dice si puede darse el caso de que efectivamente *c* cause *e* y también *b**. ¿Cuál sería el problema de que lo hiciera? Si sumamos esto a la idea de que todo suceso tiene una causa, entonces todo suceso mereológicamente sobreveniente tendría también una causa; esto presupondría una sobredeterminación global mixta (causal y sobreveniente) de todo suceso macrofísico, lo cual es muy implausible -e inelegante, diría Sabatés-. No parece ocioso preguntarse si es menos implausible la idea de que existe una sobredeterminación mixta de toda una amplia clase de sucesos, que la idea de una sobredeterminación causal. El caso contrario implica aceptar que los sucesos macroscópicos son no causados. Cuando Sabatés señala las dificultades que trae aparejadas el principio (PED), observa lo siguiente

Consideremos de nuevo: el cambio en la concentración de iones de hidrógeno en una solución es un determinante suficiente para el cambio en el valor del Ph de la solución. ¿Es entonces el cambio en el Ph de la solución no causado? Si tenemos la convicción plausible de que todo macro-suceso depende de sucesos microfísicos, entonces *todo* macro-suceso (incluyendo químicos y biológicos) sería no causado (1996, p. 106. Las cursivas son de Sabatés).

En esta continuación del ejemplo anterior, parece inclinarse por la idea de que los sucesos macrofísicos también serían causados (además de sobrevenientes), ya que

la exclusión de esta posibilidad le resulta visiblemente inaceptable. No obstante, me parece que sus análisis no esclarecen algunas cuestiones fundamentales: la primera, relativa a la relación entre el suceso sobreviniente, su base de superveniencia y la causa de los cambios en esta base; la segunda, derivada de la anterior, implica resolver de que manera un suceso sobreviniente puede a la vez ser causado (si se acepta esta posibilidad) y argumentar por qué una sobredeterminación global no implica una concepción ontológicamente implausible de la realidad.

Aceptar sucesos no causados puede resultar muy inquietante. Pero es necesario destacar en primer lugar que no implica que existan sucesos que ocurren de manera anárquica. En segundo lugar, el negar esta posibilidad implica dos dificultades de magnitud no despreciable: una, el explicar por qué la sobredeterminación global mencionada puede constituir una pintura aceptable de la realidad, y la segunda mostrar cómo un suceso que sobreviene a otro es además causado por un tercero -y no se quiere que se pruebe su *relevancia causal*, sino su *eficacia causal*, que es algo muy distinto-. Quizás para alguien que no sostenga un cerrado causalismo explicativo, una pintura de la realidad que la considere estructurada en base a una combinación de determinantes causales y de superveniencia no resulte tan inaceptable como una que sostenga una sobredeterminación global de la misma.

La relación de superveniencia, por otra parte, ha sido considerada en general como una relación estructurante distinta y complementaria de la causalidad. Esto surge claramente en el siguiente pasaje de Kim

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, p. 154, citado en Pérez, 1996).

Por otra parte, Kim no parece sentirse demasiado incómodo con la idea de un mundo estructurado de forma tal que en el mismo algunos sucesos no tienen causa

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. Esto, no obstante, requiere ser reconsiderado. Parece haber relaciones de dependencia entre sucesos que no son causales, y, como argumentaré, *el determinismo universal puede ser verdadero aún cuando no todo suceso tenga una causa* (1974, p. 22. Las cursivas son mías).

Creo que, en líneas generales, cualquier teórico dispuesto a aceptar el pluralismo explicativo se sentiría mas conforme con la idea de que las relaciones de determinación complementarias a la causalidad son completamente independientes de aquella, que con la suposición contraria. Esta última implicaría, de alguna forma, que el "cemento" básico de la realidad lo constituye la relación causal, la cual sería complementada, en relación a cierta clase de hechos, por una relación de determinación secundaria, como la superveniencia. Esto llevaría, además, a que las explicaciones de cualquier suceso serían, al menos parcialmente, causales.

Considero, en síntesis, que el aceptar la superveniencia como una relación estructurante de la realidad complementaria a la causalidad conduce naturalmente a alguna clase de exclusión de determinantes,

y a un correspondiente principio de exclusión explicativa.²⁶

²⁶ Supongamos que aceptamos que la dependencia de Cambridge constituye un tipo de relación estructurante en el mundo. Sócrates muere como efecto de haber ingerido cicuta. En el momento en que Sócrates muere, Xantipa se transforma en viuda. Como Kim sostiene, hay elementos para pensar que la relación entre los dos tipos de sucesos no es de identidad, ni tampoco causal (*cf.* Kim, 1974). ¿Es la ingestión de la cicuta por parte de Sócrates la causa de que Xantipa se transforme en viuda? No; sólo es la causa de la muerte de Sócrates. Lo que determina que Xantipa se transforme en viuda es el hecho de que Sócrates muera. La cicuta es la causa (y sólo una de las posibles) de que Sócrates muera; la muerte de Sócrates es el determinante necesario y suficiente para la viudez de Xantipa. En este caso, tendríamos nuevamente una estructura de sucesos en el cual existe una relación causal (ingestión de cicuta-muerte de Sócrates), y una relación de determinación no causal, la dependencia de Cambridge (muerte de Sócrates-viudez de Xantipa). Si la dependencia de Cambridge es un determinante suficiente para la viudez de Xantipa, entonces ningún otro determinante es necesario para su producción. La presencia del mismo excluye a cualquier otro tipo de relación en la producción del suceso.

III. SUPUESTOS E IMPLICACIONES EPISTEMOLÓGICAS DEL

PRINCIPIO DE EXCLUSIÓN: UNA CRÍTICA

Hemos visto en el capítulo I que el planteo del problema de la exclusión causal-explicativa de Kim, a diferencia de sus antecesores (los de Malcolm y Goldman), y de otros planteos contemporáneos vinculados mas o menos cercanamente (los de Block, Drestke, y Jackson y Pettit), tiene un alcance mas amplio y consecuencias epistemológicas mas profundas. Como se habrá advertido, los alcances del problema de la exclusión, tanto causal como explicativo, exceden ampliamente la discusión del problema de la causación mental, si bien éste constituye un caso privilegiado de análisis. Pero el análisis de Kim, según creo, tiene una fuerza argumentativa dispar: mientras que en el plano estrictamente causal, y en especial en lo relativo a la causación mental, sus tesis están convincentemente defendidas, no ocurre lo mismo en lo referente a la defensa del principio general de exclusión explicativa.

En este capítulo me propongo examinar los presupuestos epistemológicos del principio general de exclusión que Kim formula y algunas de sus presuntas consecuencias. En primer término, intentaré mostrar que no todas las teorías de la explicación son compatibles de manera irrestricta con el principio de exclusión explicativa. En segundo lugar, analizaré los fundamentos de la aceptación del principio de exclusión en sus dos versiones, y señalaré lo que considero son insuficiencias en la argumentación tendiente a probar la aceptabilidad de la versión epistemológica del mismo. Por último, expondré un argumento destinado a demostrar que la fuerza que el principio de exclusión carece de la fuerza que Kim le atribuye.

1. Consecuencias epistemológicas en el planteo de Malcolm

Que la exclusión entre explicaciones neurofisiológicas y explicaciones propositivas tiene importantes consecuencias gnoseológicas era algo que ya se ponía de manifiesto en el artículo originario de Malcolm, el cual advertía acerca de ciertos problemas que el mecanicismo planteaba. Si bien afirmaba que ésta era una doctrina concebible en el sentido de que no era lógicamente autocontradictoria, admitía que había aspectos en los cuales no constituía una posición 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 para sostener que el mecanicismo es verdadero. Esta es una segunda paradoja,

según Malcolm, del mecanicismo: éste no puede ser sostenido sobre bases racionales.²⁷

Los dos aspectos en los cuales el mecanicismo no es concebible no implican, obviamente, que sea falso. Pero es evidente que conduce a serias paradojas. Para evitarlas, dice Malcolm, una alternativa es recurrir al solipsismo, y negar que el mecanicismo sea universalmente verdadero.²⁸

Tal vez podría discutirse que el mecanicismo, tal como lo entendió Malcolm, conduzca directamente a un mundo en el cual el libre albedrío resulta reemplazado por el determinismo de los mecanismos neurofisiológicos. Pero no me detendré aquí a analizar estas implicancias del problema. Es evidente que la aceptación de estas consecuencias del mecanicismo traería aparejadas extraordinarias modificaciones en nuestra concepción de la acción humana, de los procesos sociales, de los problemas éticos, etcétera, para no hablar de una ontología general. Dada la extrema generalidad que el problema adquiriría, no parece una línea de defensa adecuada de las explicaciones propositivas el análisis de las modificaciones que esta visión impondría a nuestra concepción del problema mente-cuerpo, del cual la relación explicaciones mecanicistas-explicaciones propositivas es una parte. De hecho, ni la respuesta de Goldman ni los análisis posteriores de Kim incluyen una reflexión sobre estas presuntas consecuencias del planteo mecanicista.

²⁷ En Churchland (1981) se encuentra una breve discusión de un intento de *reductio ad absurdum* de los postulados del materialismo eliminativo que guarda cierta relación con las paradojas mencionadas por Malcolm. Debe advertirse que la posición a la que Malcolm se ve conducido al aceptar que la proposición "el mecanicismo es verdadero" es verdadera aunque no exista ningún hablante que pueda pronunciarla con sentido, parece requerir la aceptación de una suerte de "Mundo 3" popperiano, en el cual las proposiciones pueden ser verdaderas aunque desapareciera cualquier organismo capaz de afirmarlas.

²⁸ No deja de ser sugestivo que el propio Malcolm no se mostrara convencido de la corrección de su propio análisis con respecto al punto crucial del mismo: la posibilidad de construir explicaciones neurofisiológicas y propositivas con respecto a una y la misma secuencia de movimientos. Tampoco se mostraba enteramente convencido de la corrección de una visión que hiciera perder a una secuencia de movimientos su aspecto de acción. Transcribo su ilustrativo comentario final: "Quizás la publicación del presente artículo estará justificada si provoca una defensa verdaderamente convincente de la compatibilidad de las dos formas de explicación" (1968, p.72).

2. Los alcances epistemológicos del principio de exclusión

Las consideraciones epistemológicas de Kim tienen, como se pondrá de manifiesto, un carácter más esquemático y tentativo²⁹ que sus análisis respecto de los problemas de la causación mental; no obstante, el examen de las mismas resultará ilustrativo de las dificultades que implica la aplicación del principio de exclusión explicativa. Este carácter esquemático de las consideraciones epistemológicas se manifiesta marcadamente en sus observaciones acerca de la concepción kuhniana de la ciencia. Si bien este análisis constituye un aspecto de importancia marginal en el planteo de Kim, lo examinaré a fin de comenzar a desarrollar a partir de un caso concreto lo que constituirá una crítica de mayor importancia: que la fuerza real del principio de exclusión explicativa dista de ser la que le atribuye.³⁰

En una de las "aplicaciones" de su teoría acerca de la exclusión explicativa, expuesta en su (1989a), Kim considera que la misma puede servir para explicar un aspecto "de otro modo enigmático" de la teoría de la ciencia de Kuhn, en particular, acerca 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 propuestas a los mismos. No obstante, dice Kim, a pesar de que Kuhn remarcó que los distintos paradigmas no comparten los problemas ni, en

²⁹ Para despejar la sospecha de que este carácter tentativo y esquemático en la argumentación en favor del principio general de exclusión explicativa sea producto de un análisis sólo de los primeros trabajos de Kim en los que se refiere al problema, señalaré que su (1989a) está reimpresso en su (1993b), sin modificaciones. En el Prefacio de este último libro, Kim observa que sobre algunos de los problemas expuestos no mantiene la misma opinión, y formula unas pocas observaciones acerca de su pensamiento al respecto en ese momento. Sin embargo, no señala entre ellos a las tesis defendidas en su (1989a). No queda otra alternativa, entonces, que someter a crítica estos argumentos en el estado en el cual los expuso Kim por vez primera.

³⁰ En el capítulo III expondré en detalle una continuación de esta crítica pero referida al caso de la psicología *folk*.

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 a los restantes paradigmas en competencia.

Creo que el argumento de Kim es notoriamente inconcluyente con respecto a su pretensión de que el principio de exclusión explicativa permite esclarecer este aspecto. En primer lugar, 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 está 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 desde el mismo; sólo su falta de ingenio, dice 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? Creo que es 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 mas 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 rivales, sólo podría tener, desde su punto de vista, efectos deletéreos sobre la organización social de los científicos y, consiguientemente, sobre la investigación normal.

A partir de todas estas consideraciones creo que se pone de manifiesto que esta interpretación de la necesidad de un único paradigma se ajusta mejor que la de Kim al espíritu de la obra de Kuhn. Me 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.

3. ¿Es el principio de exclusión neutral con respecto a las teorías de la explicación?

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 cuidadoso 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 (...) No obstante, 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 mas) explicaciones para un mismo hecho. Sugiere que esta situación se produce debido a que las explicaciones múltiples de un mismo *explanandum* son contraproducentes en relación al *desideratum* de simplicidad y unificación entre los ítemes 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.

Nótese que, en este planteo, Kim se basa explícitamente en la teoría de la explicación desarrollada por M. Friedman y posteriormente por P. Kitcher. Ahora bien, esta teoría de la explicación es caracterizada por Kim en su (1994b) como "internalista explicativa".³¹ Un internalista explicativo (como es el caso de Hempel, según Kim) puede aceptar una violación del principio de exclusión en ciertos casos, como por ejemplo ante presuntos casos de

³¹ No se trata de un anacronismo. En su (1988) Kim caracterizaba al modelo hempeliano de explicación como "irrealista explicativo", o también "internalista explicativo". Si bien en ese artículo no se refiere a la teoría de la explicación como unificación, no cabe dudas acerca de que esta última es heredera del modelo hempeliano, por lo que las consideraciones sobre su presunto internalismo hechas con referencia a éste son igualmente válidas para la primera.

sobredeterminación causal.³² Como he observado con anterioridad, la posición de Kim con respecto a los posibles casos de sobredeterminación explicativa no es clara.³³ Pero, como señalé, Kim sugiere que el principio de exclusión explicativa podría aplicarse aún ante genuinos casos de sobredeterminación. Con esta afirmación, Kim parece asignarle al principio un carácter más taxativo que el de una simple "restricción general plausible", la cual podría admitir, sin perder fuerza, excepciones aisladas.

Un teórico del modelo nomológico-deductivo puede admitir válidamente dos explicaciones completas e independientes de un mismo *explanandum* en los casos de sobredeterminación, por lo cual el principio general de exclusión explicativa no es neutral con respecto a cualquier teoría de la explicación. Dicho en otros términos, el principio de exclusión explicativa no tiene la misma fuerza en todas las teorías de la explicación. Podría objetarse que los (presuntos) casos de sobredeterminación constituyen una limitación muy marginal para las teorías de la explicación, y que esto implica una cualificación muy menor del planteo de Kim. Sin embargo, hay autores que han considerado que existen genuinos casos de sobredeterminación con los cuales las teorías de la explicación deben confrontar (al respecto, *cfr.* Loeb, 1974).

Quisiera hacer aquí una observación adicional. Si bien una teoría internalista de la explicación puede admitir más de una explicación completa e independiente de un mismo hecho en los casos de sobredeterminación, esto no puede constituir, de manera alguna, una solución para el problema de las propiedades mentales. En primer lugar, porque esto supondría una sobredeterminación global de la conducta humana, lo cual nos sitúa en una ontología implausible. En

³² En rigor, Hempel no habla acerca de sobredeterminación causal, sino solamente de sobredeterminación explicativa. En mi (1998) analizo la posición de Hempel al respecto (*cfr.* Hempel, 1965, pp. 411-12).

³³ *Supra*, p. 20n.

segundo lugar, porque la sobredeterminación misma de las acciones es imposible: como vimos en el capítulo anterior, la acción de una propiedad M para causar una propiedad P está prohibida por el principio de clausura causal del mundo físico. Y dado que no puede haber una propiedad M y una propiedad P actuando como causas separadas e independientes en la producción de otra propiedad (sea M o P), no puede haber sobredeterminación explicativa.

4. El principio de exclusión y el rechazo del realismo explicativo

He observado en el capítulo I que el principio de exclusión, en su versión epistemológica, no requiere de la aceptación del realismo explicativo (y tampoco, por lo tanto, de la aceptación previa del principio metafísico). ¿Cuáles son, entonces, los argumentos para aceptar este principio? En primer lugar, la idea ya expuesta de que encontramos algo disonante y perturbador al poseer dos explicaciones independientes del mismo hecho, lo cual se produce porque ambas explicaciones, tomadas conjuntamente, pueden producir algún tipo de incoherencia en nuestro sistema de creencias. La frase de Kim "Anteriormente intenté explicar por qué esto podría ser así sobre la base de la concepción según la cual el rol epistémico primario de la explicación es la simplificación y la unificación de nuestro sistema de creencias" (1989a, p.258), ¿implica que la teoría de la explicación como unificación es igualmente la base para sostener al principio epistemológico? ¿O es sólo una mención de la que considera la posición plausible *cuando se acepta el realismo explicativo*? Por otra parte, la mención a la teoría coherentista de la justificación, ¿es un argumento para sostener el principio epistemológico, o es sólo una ilustración de una manera de entender el tipo de incoherencia suscitada por la coexistencia de dos (o

mas) explicaciones? 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 la misma y el irrealismo explicativo³⁴ se combinan adecuadamente entre sí.

Veamos primero que es lo que sostiene un irrealista explicativo, bajo la perspectiva de Kim:

Irrealismo explicativo (...) sería la opinión de que la relación de ser un *explanans* para, como lo que relaciona C y E dentro de nuestro *corpus* epistémico, no está y no necesita estar "basada" en ninguna relación objetiva entre los eventos *c* y *e*. Es solamente una cuestión de alguna relación "interna" entre ítemes 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 eventos representados como estando relacionados por una relación explicativa (Kim, 1988, p. 226).³⁵

Kim agrega 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 misma, teniendo en cuenta la verdad de sus enunciados componentes.³⁶

³⁴ En su (1987) Kim denomina a esta posición "irrealismo explicativo", si bien también sugiere que "internalismo explicativo" sería tal vez una expresión mas adecuada. En su (1994b) sólo emplea la segunda expresión para referirse a esta posición.

³⁵ Hasta donde he podido indagar, el ejemplo privilegiado que Kim aporta como demostrativo de una posición internalista explicativa es el modelo nomológico-deductivo de Hempel (si bien también considera como internalistas a otros teóricos de la explicación, como Friedmann y Kitcher).

³⁶ Nótese que el internalismo explicativo 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 buena 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.

Las tesis características de las teorías coherentistas de la justificación, según Haack, son las que afirman que

la justificación es exclusivamente un asunto de relaciones entre creencias, y que 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 (Haack, 1995, p. 17).
37

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 las proposiciones componentes de la misma 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 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, creo, donde surgen 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. Tal como se expuso en el capítulo II,

³⁷ 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 mas general de esta posición.

Sabatés considera 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 mas 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 me parecen algo esquemáticas,³⁸ pero pueden proporcionar la base para un argumento en contra de la aceptación del principio de

³⁸ Los problemas que surgen de este enfoque se producen, a mi juicio, debido a la insuficiente fundamentación con la que Sabatés descarta a la exclusión a nivel ontológico (*supra*, p. 42 y ss.). Otros problemas pueden ser suscitados por la deficiente elucidación de los conceptos en juego. En particular, ¿qué debemos entender por "comprensión" del fenómeno, y cómo aumentan nuestra comprensión del mismo dos explicaciones rivales?

exclusión explicativa epistemológico sin admitir el principio metafísico.

En primer lugar, se debe hacer notar que, si los resultados expuestos en el capítulo II son correctos, si existen bases ontológicas para la exclusión explicativa. No obstante, 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 mas 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).

Tratemos ahora de determinar en cuales 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 que en ella se hagan referencias 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*. 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 (son de alguna manera "incommensurables"), 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 mas información acerca de este aspecto del mundo (y,

presumiblemente, acerca de otros), ¿por qué no conservar ambas explicaciones (y, consecuentemente, ambas teorías)? No encuentro una buena razón 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 estoy afirmando que siempre es preferible un estado de menor simplicidad y unidad en nuestro conocimiento, compensado por un aumento de teorías explicativas; sólo afirmo que, *en algunos casos*, la pérdida de simplicidad y unidad puede verse compensada por la mayor información proporcionada por teorías explicativas rivales.³⁹ Si, en algunos 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.

Si se objetara que en realidad no existe tal aumento de conocimiento, porque las explicaciones ofrecidas por una de las teorías rivales son falsas, se estaría sosteniendo en realidad que las entidades explicativas postuladas por la teoría no existen, o que las leyes y generalizaciones que postula son falsas; pero si este fuera el caso, las teorías explicativas propuestas serían inadecuadas en un

³⁹ Deseo subrayar el hecho de que, a diferencia de Sabatés, no sostengo que las dos explicaciones rivales acerca del mismo *explanandum* aumenten nuestra *comprensión* del mismo. Dos razones me llevan a rechazar esta alternativa: en primer lugar, por las conocidas dificultades de lograr una caracterización mínimamente plausible del concepto de "comprensión"; en segundo lugar, porque a pesar de esta dificultad creo 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 parece adecuado sostener que tales explicaciones puedan aumentar nuestra comprensión del *explanandum*.

sentido realista. Para descartar el supuesto de la compensación de la pérdida de simplicidad y unidad por medio de la ganancia de conocimiento, se debería entonces apelar a consideraciones realistas. Dicho en otros términos, esto último implicaría que una de las dos explicaciones no refleja lo que ocurre en el mundo, por lo cual no constituye una ganancia epistémica legítima. La aceptación irrestricta del principio epistemológico, en síntesis, requeriría de la aceptación previa del principio metafísico.

5. El principio de exclusión y la elección de teorías

Como hemos visto, el principio metafísico de exclusión explicativa se basa claramente 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.⁴⁰ El principio metafísico supone entonces como premisas, la aceptación de la teoría de la explicación como unificación y del realismo explicativo. Si aceptamos estas dos premisas, esto 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 es, para Kim, un caso especial de la "navaja de Occam", o principio de simplicidad: es una "regla específica que concierne a una

⁴⁰ O, como hemos visto, que se trate de un caso de sobredeterminación.

manera importante en que se gana simplicidad en cuestiones explicativas y explica por qué es deseable esta forma de simplicidad" (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 no se cuenta con 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* vs. 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 mas explicativa, la única manera en que puede sobrevivir es siendo reducida de manera conservadora a esta última*" (1989a, p. 261. Las cursivas son de Kim). Es decir, una de las teorías debe desaparecer. 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 mas de una explicación completa e independiente de un mismo suceso".⁴¹ 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

⁴¹ 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 me parece pertinente, creo que 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 hace referencia a la manera en que una teoría debe relacionarse con otra para no ser eliminada, cuestión sobre la que el primer principio no afirma nada.

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 resulta fácil determinar cual es el carácter que Kim le asigna. Si bien, como se comentó anteriormente, 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. Creo que son posibles dos respuestas:

1. 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.⁴²
2. Es una *regla efectiva* que permite decidir el reemplazo de una teoría por otra en casos reales de elección. En este caso, el mayor poder explicativo de una teoría con respecto a otra es 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.

⁴² Al formular esta observación no estoy exponiendo mi propio punto de vista. Simplemente estoy señalando que, como Kim mismo 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 la misma puede acreditar. Por esta 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.

Si el principio (EET) fuese considerado como teniendo el alcance establecido en la afirmación 1, creo que sería difícil negar su plausibilidad (al punto que resultaría trivialmente verdadero). Si el alcance que se le atribuyera fuese el establecido por la afirmación 2, la situación resultaría bastante mas compleja. Podríamos preguntar, en primer lugar, si el principio (EET), bajo la interpretación 2, 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, creo, a situar el problema en un marco mas amplio, esto es, el de cuales son los criterios que permiten decidir racionalmente el reemplazo de una teoría por otra. La pregunta pertinente sería: ¿es el principio de exclusión 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 . Luego, para el mismo fenómeno x surge la teoría explicativa Z , 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 ambas son incompletas y pueden ser unificadas sin eliminación)? En primer lugar, que ambas teorías se *superpongan explicativamente de una manera completa*, esto es, todo lo que explica la primera teoría debe ser explicado por la segunda. Si hubiera alguna clase de pérdida, el criterio una ganancia epistémica de unificación de nuestro conocimiento a costa de una pérdida de poder explicativo no tendría suficiente fuerza. En segundo lugar, como Kim sostiene, que una de las teorías explique mejor que la otra

los "casos interesantes".⁴³ 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 aspectos, nuevamente el criterio de ganancia epistémica por medio de la unificación 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 paradigmáticamente señalados como de unificación exitosa (como 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 aspectos relevantes, es decir, no solamente tiene mayor capacidad explicativa, sino también capacidad predictiva, precisión, etc.

Podría objetarse que estas consideraciones sólo tienen alguna fuerza argumentativa si se admite que valores epistémicos tales como la simplicidad, el poder heurístico, la compatibilidad con otras teorías y otros campos, etc., son tan importante como el apoyo empírico (el cual incluiría el alcance explicativo, la capacidad predictiva, la contrastabilidad, etc.). Pero creo que 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, creo suficiente hacer notar que los mencionados procesos implican una complejidad considerable que el análisis de

⁴³ El concepto de "mejor desempeño" no es elucidado por Kim, aunque puede intuirse que su análisis conduciría a complicaciones adicionales.

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. Creo que 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. Creo además que en este esfuerzo debería proporcionar 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.

No carece de interés señalar, en relación al análisis anterior, que el modelo de "reducción eliminativa" al que Kim apela parece implicar consideraciones adicionales, además de 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, 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 John Kemeny y Paul

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 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, dado que la reducción implica 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. Las cursivas son 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 mas 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 mas simple sea reemplazada por una mas 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 mas fuerte, parecería razonable permitir alguna complejidad adicional.⁴⁴ Resultaría satisfactorio, entonces, que alguna 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,⁴⁵ 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*. Mas 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.

Creo haber mostrado, a partir de las observaciones precedentes, 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

⁴⁴ Es necesario hacer notar que la simplicidad que se obtiene con la disminución del número de términos teóricos que nos vemos obligados a aceptar al adoptar la teoría reductora no implica, *per se*, una mejora en nuestra situación epistémica. Kim señala, en relación al problema de los vínculos entre simplicidad y comprensión, que los sistemas mas 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).

⁴⁵ Kemeny y Oppenheim advierten en una nota que el concepto de simplicidad necesita de "considerable estudio", si bien existen algunos resultados parciales.

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 a 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, aún cuando la comunidad científica ya no la considere como adecuada de acuerdo a los parámetros de aceptación racional vigentes. Esta situación es posible aún cuando la teoría eliminada sea considerada refutada; en este caso, la simplicidad de la misma para el diseño de tecnologías eficientes compensa en una medida considerable sus deficiencias como marco para la comprensión del mundo. Un ejemplo notable de esta situación es nuevamente proporcionado por la teoría de Newton. Si bien la teoría es considerada falsa, no hay duda de que su aplicación con fines tecnológicos en relación a ciertos sistemas físicos es notablemente más sencilla que la teoría de la relatividad, por lo que se la sigue utilizando para estos propósitos. Esto marca una diferencia notable entre esta teoría y la del flogisto: la segunda desapareció para siempre; su eliminación fue, por lo tanto, absoluta. En un cierto sentido, evidentemente más débil que el sostener que la teoría sigue teniendo vigencia plena para la comunidad

científica, pero no completamente irrelevante, puede decirse que la teoría no queda completamente eliminada del campo del conocimiento. Todo esto implica 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.

III. EL CASO DE LA *FOLK PSYCHOLOGY*

En este capítulo me propongo estudiar el que podríamos denominar "caso testigo" para la viabilidad de un principio de exclusión para las explicaciones psicológicas en los términos en los que es afirmado por Kim. El análisis de este caso tiene importancia por dos razones: la primera, que es el propio Kim quien presenta a la *folk psychology*⁴⁶ como el candidato paradigmático a la eliminación por medio de la aplicación del principio de exclusión explicativa; si en el caso que él mismo alega la aplicación del principio no tuviera la fuerza que reclama, su pretensión acerca de los alcances del mismo se vería considerablemente debilitada. La segunda, que este caso constituye uno de especial importancia para la existencia de una ciencia psicológica independiente tal como se la desarrolla actualmente. Si bien el marco intencional deseo-creencia no tiene la misma importancia en todos los dominios de investigación contemporáneos, no cabe duda de que constituye el "sustrato" conceptual de diversas teorías y áreas de investigación de importancia.⁴⁷ Por esta razón, su potencial eliminación constituye una cuestión de relevancia para evaluar los alcances del principio de exclusión explicativa.

Con este propósito, examinaré en primer lugar la concepción clásica de materialismo eliminativo expuesta por Paul Churchland (1981) quien, además de uno de los principales teóricos del eliminativismo contemporáneo, es

⁴⁶ No existe, que yo sepa, una traducción unánimemente aceptada para la expresión inglesa "*folk psychology*". Algunas de las utilizadas son "psicología ordinaria", "psicología de sentido común", "psicología popular", y también "psicología *folk*", entre otras. En lo que sigue, utilizaré la expresión "psicología de sentido común" (PSC).

⁴⁷ Un ejemplo particularmente notable lo constituye la llamada "teoría de la acción razonada", de I. Ajzen y M. Fischbein. Esta es una teoría general de la conducta humana que liga la realización de una determinada conducta a la presencia de ciertas intenciones, las cuales son a su vez función de determinadas creencias, normas y actitudes hacia la realización de esa conducta. La presencia del marco conceptual de la PSC se aprecia nítidamente en esta descripción esquemática. Para una discusión acerca de la relación entre psicología de sentido común y ciencia cognitiva, *cfr.* Engel (1988).

el autor que Kim analiza al presentar las consecuencias de la aplicación del principio de exclusión explicativa. En segundo lugar, comentaré las defensas de la PF expuestas por Horgan y Woodward (1985) y por Jackson y Pettit (1990b). No es mi propósito argumentar en favor o en contra de la existencia independiente de la PSC, si bien haré algunas observaciones aisladas con respecto a algunos argumentos particulares. Mi objetivo general consistirá en exponer y evaluar los argumentos a favor y en contra de la existencia de la PSC, para determinar en que medida se hace una utilización explícita o implícita de algún principio de exclusión explicativa, o se apela a consideraciones epistemológicas mas generales. Es importante señalar que los argumentos que se expondrán en defensa de la PSC aceptan que la misma constituye una teoría empírica. En caso de que no lo hicieran, apelando a una concepción de la PSC que privilegie un presunto carácter normativo de la misma para la evaluación de las acciones y la formación de intenciones, esto la ubicaría mas allá del alcance de cualquier principio de exclusión explicativa y de cualquier criterio evaluativo de adecuación empírica.

Si mi análisis es correcto, pretendo mostrar que, en concordancia con las conclusiones expuestas en el capítulo precedente, el principio de exclusión explicativa expuesto por Kim no tiene la fuerza que él pretende, y que, si bien constituye una "restricción general plausible" para las explicaciones, no es un criterio de juicio único y decisivo para la evaluación de teorías en competencia.

1. El principio de exclusión como fundamento de la eliminación

Como he mencionado en el capítulo I, y analizado en el III, Kim asigna al principio de exclusión cierto valor en la elección de teorías explicativas rivales. Kim comienza

este planteo con un análisis del modelo de reducción debido a Kemeny y Oppenheim.⁴⁸ El abandono de la teoría reducida de acuerdo con este modelo tiene, según Kim, drásticas consecuencias: sus entidades teóricas características, sus propiedades y sucesos ya no se reconocen como "reales"; tal como el flogisto, desaparecen cuando la teoría que las postula es descartada.

Las consideraciones mencionadas, sostiene Kim, ayudan a esclarecer el debate sobre el futuro de la denominada "psicología ordinaria" y su relación con una ciencia cognitiva desarrollada. Muchos filósofos, prosigue, han planteado dudas acerca de la realidad de esta psicología y de la existencia real de estados tales como creencias, intenciones y deseos. Una ciencia cognitiva desarrollada, según ellos, reemplazará a la psicología ordinaria de forma tal que en el futuro resultará racional pensar que nunca han existido los mencionados estados intencionales. Uno de los mas notables defensores de esta posición ha sido Paul Churchland, quien se basa en dos consideraciones, según Kim, para argumentar en favor de la eliminación de la psicología ordinaria. El primero de estos argumentos (que describiré en el siguiente apartado) consiste en señalar los serios fracasos explicativos que afronta la psicología ordinaria; el segundo, la altamente improbable reducción conservadora de la misma a una teoría científica sistemática acerca de lo mental. ¿Por qué, pregunta Kim, Churchland concluye en que ambas psicologías no pueden coexistir, si no se excluyen lógicamente o empíricamente? Esto puede explicarse, en su opinión, por la aplicación del principio de exclusión explicativa. Debido a que la psicología ordinaria y una ciencia desarrollada acerca de lo mental intentan proporcionar explicaciones independientes y completas para un mismo dominio de hechos, y dado que no puede efectuarse ninguna reducción, de

⁴⁸ *Supra*, p. 73.

acuerdo con el principio de exclusión explicativa una de las dos debe desaparecer.

Kim agrega que, en su opinión, la manera correcta de salvar la psicología ordinaria es dejar de considerarla una teoría empírica cuyo objetivo es proporcionar explicaciones y predicciones causales basadas en leyes. Resultaría preferible, en su opinión, centrarse en su rol normativo en la evaluación de acciones y la formación de intenciones y decisiones. En este último caso, dejaría de competir con una ciencia cognitiva desarrollada; en caso de competencia, lo mejor que cabría esperar sería su absorción reductiva por una teoría mas sistemática y mejor fundada.

En el resto de este capítulo mostraré en primer lugar que los argumentos eliminativistas de Churchland no se basan únicamente en consideraciones explicativas, sino que propone un marco mas amplio para la consideración de los méritos de teorías rivales; en segundo lugar, señalaré que tampoco se da una aplicación tácita o explícita del principio de exclusión en algunas de las principales defensas de la PSC. Por último, expondré algunas conclusiones que vinculan el tema tratado en este capítulo con resultados expuestos en los capítulos anteriores.

2. Churchland y los argumentos en favor del eliminativismo

Paul Churchland ha sido uno de los mas conspicuos defensores de la eliminación de la PSC, en favor de una neurociencia avanzada. La PSC, sostiene Churchland, constituye un "cuerpo integrado de saberes populares que conciernen a las relaciones legaliformes que valen entre circunstancias externas, estados internos y conducta pública" (1981, p. 45). Gracias a ella, las personas son capaces de explicar e incluso predecir con gran facilidad y éxito la conducta de sus semejantes. Usualmente, sostiene, estas explicaciones y predicciones hacen referencia a

deseos, creencias, intenciones, percepciones, etcétera, que supuestamente los agentes experimentan. Estas explicaciones y predicciones presuponen leyes (aunque sea toscas y operativas) que conectan las premisas de la explicación con la conducta externa.

Pero a pesar de este éxito saliente, Churchland considera que es posible que este marco esté radicalmente equivocado. En uno de sus escritos ha definido al eliminativismo de la siguiente forma

El materialismo eliminativo es la tesis que sostiene que nuestra concepción de sentido común acerca de los fenómenos psicológicos *constituye una teoría radicalmente falsa*, una teoría tan esencialmente defectuosa que *tanto sus principios como su ontología* serán eventualmente desplazados, mas que reducidos con fluidez, por una neurociencia completa (*ibid.*, p. 44. Las cursivas son mías).

En esta definición se encuentran los principales elementos componentes del argumento en favor del eliminativismo de Churchland: la PSC constituye una teoría empírica, y, mas aún, una teoría empírica que ha sido refutada. Su inadecuación no se limita al hecho de que algunas de sus leyes han resultado refutadas en su confrontación con la experiencia;⁴⁹ se trata de que las entidades a las que estas leyes refieren simplemente *no existen*, por lo que la teoría constituye una forma básicamente errónea de concebir tales fenómenos. El materialista eliminativo, dice Churchland, coincide con el dualista y el funcionalista en rechazar la posibilidad de la reducción de las entidades propuestas por la PSC, pero lo hace a partir de elementos de juicio muy distintos: la PSC constituye una explicación radicalmente inadecuada,

⁴⁹ Kim advierte que la afirmación de que las leyes de la PSC han sido refutadas es objeto de controversia, y que, en el mejor de los casos, no forzaría a su eliminación sino a su mejoramiento. No puedo pronunciarme acerca de la corrección de la primera parte de la afirmación de Kim, pero creo que la segunda es opinable. El hecho de que las leyes fundamentales de una teoría deban ser rechazadas o mejoradas como resultado de una refutación depende, obviamente, de la teoría de la aceptación racional de las teorías científicas que se sostenga. Un popperiano estaría de acuerdo, *en principio*, con la afirmación de Churchland.

demasiado confusa y defectuosa como para poder sobrevivir a una reducción interteórica.

El primer paso que Churchland debe dar para argumentar en favor de la eliminación es probar que la PSC constituya una teoría, afirmación que ha sido rechazada por muchos filósofos a partir de argumentos muy distintos. Admitir que la PSC es una teoría, sostiene Churchland, "proporciona una organización simple y unificadora de la mayoría de los tópicos principales de la filosofía de la mente" (*ibid.*, p.44). Estos tópicos incluyen la explicación y la predicción de la conducta, la semántica de los predicados mentales, la teoría de la acción, el problema mente-cuerpo, etcétera. Por citar sólo un ejemplo, veamos como se resuelve, según Churchland, el problema de las otras mentes dentro del marco de la PSC. La convicción intuitiva y problemática de que otras personas poseen estados mentales no es inferida deductivamente de la conducta de esas personas, ni por una extremadamente endeble generalización inductivo-analógica a partir del caso propio. Esta creencia en las otras mentes constituye, según Churchland, una *hipótesis explicativa singular* cuya función, en conjunción con las otras leyes de la PSC, es la de proporcionar explicaciones/predicciones/comprensión de la conducta de los otros individuos, y es creíble en la medida que resulte mas exitosa que otras hipótesis alternativas.

No me extenderé sobre estas consideraciones de Churchland, ya que mi objetivo no es analizar su argumento en sí mismo, sino en relación a la aplicación de algún principio de exclusión explicativa. En lo que sigue, entonces, daré por correctos los argumentos de Churchland, con respecto a que la PSC constituye una teoría.

Dado que la PSC es una teoría empírica, prosigue Churchland, puede explorarse la posibilidad de que constituya una teoría falsa. Para evaluar esta posibilidad es necesario estimar en primer lugar sus éxitos y fracasos

explicativos y predictivos; en segundo lugar, su historia, crecimiento, fertilidad y promesa de desarrollo futuro; y en tercer lugar, considerar que tipos de teorías acerca de nuestra conducta son probablemente verdaderas a la luz de lo que hemos aprendido acerca de nosotros mismos en épocas recientes.

Con respecto al primer aspecto, debe admitirse que la PSC ha disfrutado de un éxito explicativo y predictivo notable. A partir de sus categorías, somos capaces de predecir y explicar con sorprendente exactitud, las acciones de nuestros semejantes. No obstante, advierte Churchland, existe una variedad de fenómenos que no puede explicar, algunos de los cuales ni siquiera son abarcados dentro de su marco. Entre ellos se encuentran "la naturaleza y dinámica de la enfermedad mental, la facultad de la imaginación creadora o el funcionamiento de las diferencias de inteligencia entre los individuos" (*ibid.*, p.50). La PSC es incapaz asimismo de abordar los procesos de aprendizaje cuando implican un cambio conceptual importante y éste aparece en forma no lingüística o prelingüística, ya que su concepción del aprendizaje (manipulación y almacenamiento de actitudes proposicionales) falla al enfrentar el hecho de que formular, almacenar y manipular un conjunto amplio de actitudes proposicionales es algo que se aprende, entre otras capacidades cognitivas adquiridas.

Fracasos de semejante magnitud no prueban, concluye Churchland, que la PSC sea falsa, pero si aumentan marcadamente la probabilidad de que lo sea y demuestran que, en el mejor de los casos, es una teoría altamente superficial que no alcanza a penetrar adecuadamente en una realidad compleja y profunda.

Si ahora se examina la viabilidad de la PSC a la luz de su historia y promesa de desarrollo futuro, dice Churchland, la situación no parece ser mejor. Esta es

caracterizada como "una historia de retraimiento, infertilidad y decadencia" (*ibid.*, p. 54). En su dominio preferencial, el de la conducta humana, la PSC no ha avanzado visiblemente en los últimos dos o tres mil años. Para cualquier teoría, sostiene Churchland, este es un período demasiado largo de estancamiento e infertilidad, en especial si al mismo tiempo se enfrenta con una gran lista de anomalías explicativas en su propio dominio. Para usar los términos de Lakatos, concluye Churchland, la PSC constituye un programa de investigación degenerativo.⁵⁰

Churchland considera, no obstante, que "el éxito explicativo efectivo no es, por supuesto, la única dimensión en la que una teoría puede ser virtuosa o prometedora" (*ibid.*, p.52). Una teoría que se encuentra estancada o en problemas puede merecer atención debido a otros fundamentos. Entre ellos menciona a que es la única teoría que se ajusta correctamente a otras teorías que se ocupan de dominios adyacentes, o la única que parece poder reducirse a una teoría de fondo ya establecida cuyo dominio incluya el dominio de la teoría candidata a la reducción. Pero es en este tercer aspecto donde la PSC, según Churchland, muestra su desempeño mas bajo. Las categorías de la PSC no forman parte de la síntesis creciente que diversas disciplinas y teorías (teoría atómica y molecular, química orgánica, teoría de la evolución, etcétera) pueden ofrecer acerca de la constitución, desarrollo y capacidades conductuales del *homo sapiens*.

Churchland concluye que, debido a su estancamiento y problemas explicativos, no cabe esperar que las categorías de la PSC se vean reflejadas dentro del marco de una neurociencia avanzada.

⁵⁰ Creo que los argumentos de Churchland con respecto al desarrollo de la PSC no le hacen justicia. Creo que es éste uno de los aspectos donde mas se advierte la diferencia entre la misma y las teorías científicas *standard*. A diferencia de ellas, la PSC no ha sido el producto de una investigación deliberadamente orientada con propósitos explicativos y predictivos, por lo que juzgarla de acuerdo al criterio de progresividad como programa de investigación parece al menos discutible.

Resulta evidente, a la luz de la síntesis expuesta, que Churchland se sirve de una serie de criterios relativamente independientes para abogar en favor de la eliminación de la PSC. Las consideraciones explicativas no constituyen el criterio único en favor de esa eliminación, aún cuando la PSC se enfrente a una neurociencia completa, la cual presumiblemente la superaría ampliamente en cuanto a poder explicativo. Si bien el supuesto fracaso explicativo de la PSC no demuestra de manera concluyente su falsedad, Churchland considera que aumenta la probabilidad de que lo sea, y éste es un factor de peso en su evaluación. Para Kim, por el contrario, no hace falta que haya elementos de juicio suficientes para determinar que la teoría es falsa para que pueda procederse a su eliminación; mas aún, sostiene, puede haber razones para pensar que es verdadera. Por otra parte, Churchland introduce una consideración relativa a la evolución de las teorías en competencia. Esto implica que la evaluación de estas teorías no puede limitarse a los méritos que las mismas ostentan en el presente, se limiten o no a la capacidad explicativa, sino que es necesario ampliar la base evaluativa a su desarrollo pasado y promesa futura (en términos lakatosianos, se trataría de determinar en qué medida nos enfrentamos a programas de investigación progresivos o degenerativos). Por último, admite un criterio que podría denominarse "compatibilidad intracientífica" (o "interteórica"), ya que reclama que las teorías puedan ser compatibles con teorías aceptadas en dominios cercanos. Estos dos últimos criterios son mas amplios que la comparación de la capacidad explicativa, al cual sin duda Churchland no considera que puedan reducirse. El principio de exclusión explicativa, por lo tanto, no juega ningún rol en la defensa del eliminativismo.

3. Una defensa epistemológica de la PSC

Horgan y Woodward (1985) realizan una defensa epistemológica de la PSC contra los ataques eliminativistas.⁵¹ En líneas generales, podría decirse que, aceptando básicamente los criterios de evaluación de la PSC propuestos por Churchland, arriban a conclusiones simétricamente opuestas.

Como primer paso, Horgan y Woodward describen el análisis de Churchland como basado en tres consideraciones: a) los notables fracasos explicativos que afronta la PSC; b) el estancamiento que se advierte en la misma en los últimos veinticinco siglos; y c) su altamente improbable reductibilidad a una neurociencia desarrollada. Sostienen que si bien la irreductibilidad es la consideración principal, se ve reforzada por los restantes argumentos.

Con respecto al primer argumento, Horgan y Woodward señalan dos aspectos en los que, sostienen, resulta claramente deficitario. En primer lugar, si bien admiten que la PSC aparentemente no ha hecho un gran aporte a nuestra comprensión de fenómenos tales como la imaginación creadora o la construcción de imágenes visuales en tres dimensiones, advierten que las teorías basadas en conceptos derivados de la PSC tienen mucho que decir acerca de los mismos. Por ejemplo, las teorías de la atribución y de la disonancia cognitiva son notables casos de esta conexión entre PSC y una psicología científica desarrollada. Sostener que estas teorías no son explicativas, concluyen, es altamente implausible, y afirmarlo requeriría de una argumentación empírica detallada que Churchland no provee. Por otra parte, afirman que Churchland impone *a priori* la exigencia de que una teoría psicológica exitosa debe

⁵¹ Si bien sólo una sección del artículo de Horgan y Woodward está dedicado a discutir los argumentos de Churchland (ya que destinan una buena parte del mismo a analizar los argumentos eliminativistas de S. Stich), me centraré sólo en la discusión de los primeros, ya que son los únicos relevantes para el examen de los alcances del principio de exclusión.

proveer un enfoque adecuado y de una manera unificada para un rango preestablecido de fenómenos, y que esta exigencia debe ser considerada con escepticismo y prudencia. Para el caso de la PSC, no hay una buena razón *a priori* para esperar que la misma, diseñada primariamente para explicar acciones humanas comunes en términos de creencias, deseos, etcétera, deba proveer una comprensión adecuada de fenómenos tales como la percepción visual, el sueño, o complicados movimientos musculares.

El argumento de Churchland relativo al presunto estancamiento de la PSC, afirman Horgan y Woodward, también es claramente inconcluyente. Es al menos discutible, prosiguen, que la PSC no haya cambiado significativamente y se haya mantenido estancada en los últimos veinticinco siglos. Un posible ejemplo de su carácter empíricamente progresivo lo constituye la mayor disposición, en la cultura contemporánea, de apelar a creencias y motivaciones inconcientes.⁵² Un segundo contraejemplo al carácter empíricamente no progresivo lo constituye el hecho de que diversas teorías cognitivas contemporáneas que emplean los conceptos de deseo y creencia han conducido a nuevas y sorprendentes predicciones que han sido confirmadas por experimentos. A pesar de esto, alegan, Churchland parece sostener que el presunto carácter empíricamente no progresivo de la PSC es una buena razón para considera a cualquier teoría modelada sobre su base como falsa. Por último, y mas radicalmente, Horgan y Woodward sostienen que el criterio de "progresividad empírica" no es muy útil para evaluar una teoría como la PSC. El usuario típico de la PSC, sostienen, está interesado en aplicar una teoría preexistente para formular juicios causales acerca de casos particulares de conductas humanas, no en formular

⁵² Un interesante ejemplo de la influencia en la cultura occidental contemporánea de teorías acerca del psiquismo lo constituye la rápida difusión de los conceptos teóricos del psicoanálisis. Un notable estudio de esta difusión es la investigación pionera de S. Moscovici expuesta en su libro *El psicoanálisis, su*

generalizaciones causales nuevas. Este punto emerge claramente, afirman, cuando se advierte que mucha de la teoría implícita que está presente detrás de muchos juicios causales comunes no psicológicos ha cambiado presumiblemente poco (si es que lo ha hecho) en los últimos siglos.

Por último, el argumento mas importante de Churchland: la supuesta irreductibilidad de la PSC. Al respecto, Horgan y Woodward señalan que concuerdan en que una manera ideal (o aproximadamente ideal) de salvar la PSC sería por medio de su reducción a la ciencia natural. Y también están de acuerdo en que esto resulta improbable, dado que la PSC, en tanto que teoría milenaria, obviamente no ha sido formulada teniendo en cuenta la posibilidad de una absorción término a término por la ciencia del siglo veinte. Pero aún si la PSC no puede ser reducida a teorías de nivel inferior, y aunque estas teorías provean de una maravillosa comprensión de la naturaleza y conducta del *homo sapiens*, simplemente no se sigue que sea radicalmente falsa, o que los seres humanos no experimenten los estados intencionales que postula.

4. Una defensa funcionalista de la PSC

Jackson y Pettit (1990) ensayan otra línea de defensa de la PSC. Su argumentación general se compone de dos partes. En la primera, intentan proveer un argumento en el cual se concluya la existencia de deseos y creencias, y que no se vea amenazado de manera probable por el desarrollo de la neurociencia (lo cual es un característico argumento eliminativista). Al igual que Horgan y Woodward, admiten el postulado eliminativista de que la PSC es una teoría empírica, y por lo tanto susceptible de refutación. En

imagen y su público (1961), en el cual estudia la penetración de los términos teóricos psicoanalíticos en la

particular, consideran que constituye una teoría dentro del marco del funcionalismo de sentido común. Este argumento funcionalista, suponen, hace altamente improbable que la PSC pueda ser refutada por el avance de la neurociencia.

En la segunda parte, Jackson y Pettit intentan rebatir a otro de los argumentos típicos del eliminativismo: la imperfección predictiva de la PSC. Según ellos, la PSC exhibe las mismas dos clases de fallas predictivas que otras teorías desarrolladas en la ciencia: no formular predicciones que otra teoría realizaría correctamente, y formular predicciones erróneas. La conclusión eliminativista ante estas fallas es que la PSC será refutada por una neurociencia completa, del mismo modo general que Copérnico y Kepler refutaron a Tolomeo. A esto replican que no todos los casos de progreso científico involucran la refutación radical de lo anterior. Puede ocurrir que la nueva teoría contenga las partes relevantes de la teoría antigua. Una neurociencia completa contendrá a la PSC; junto con suficiente información acerca del entorno, nos dirá todo acerca de los roles funcionales de nuestros estados internos: cómo unos estados interactúan con otros, que es lo que hacen y que causarían, etcétera.

Jackson y Pettit consideran que pueden surgir tres objeciones ante esta tesis. La primera, a partir de la historia de la ciencia; la segunda, la que se podría denominar "objeción de la necesidad" [*objection from need*]; y la tercera, acerca del carácter retrógrado del funcionalismo. Me limitaré a las dos primeras, dado que la restante no se vincula a la evaluación de teorías empíricas. La primera, proveniente del eliminativismo, señala que existe un caso en el cual una nueva teoría con poder predictivo superior puede dejar intacto mucho del aparato teórico de la vieja teoría. El clásico ejemplo de este tipo de reducción es el de la teoría termodinámica de

los gases a la teoría cinética de los gases. Pero los eliminativistas afirman que esta excepción al desplazamiento de la vieja teoría por una nueva es la única excepción. De acuerdo con esto, una neurociencia completa dejará a la PSC intacta sólo si una identificación reductiva de deseos y creencias en términos de las nociones explicativas fundamentales de la neurociencia es posible. Jackson y Pettit concuerdan con respecto a que es una cuestión abierta si es posible una identificación reductiva de las propiedades fundamentales de la PSC en términos de las propiedades explicativas de una neurociencia completa, pero discrepan con la afirmación de que es la única manera de evitar la eliminación de la vieja teoría. Lo que importaría en el caso de un eventual reemplazo de la PSC no sería la existencia de identificaciones reductivas de las propiedades fundamentales de la vieja teoría con las propiedades de la nueva, ni la posibilidad de una reducción uniforme de la vieja teoría a la nueva; lo que importaría, por el contrario, sería que la nueva teoría contendría información que apoye las partes relevantes de la vieja teoría.

La "objeción de la necesidad" puede ser expresada en la siguiente afirmación: para decidir que es lo que se debería creer, se debe observar lo que es necesario para la ciencia; no se debe perder el tiempo en análisis conceptuales. El rol de los filósofos en la ontología es simplemente discernir qué es lo que la ciencia requiere para la explicación y predicción de los fenómenos. Esta es, según Jackson y Pettit, una posición que a veces se menciona como "naturalismo filosófico". Pero, observan, no se debería pensar que el naturalismo filosófico constituye una amenaza para la existencia de creencias y deseos. En la discusión de la segunda premisa de su argumento, creen haber encontrado excelentes razones para sostener que los estados neurofisiológicos necesarios para explicar la

conducta ocupan los roles funcionales definidos de creencias y deseos. Por lo tanto, se requieren creencias y deseos para explicar la conducta.

5. ¿Qué lugar queda para el principio de exclusión?

Como hemos visto en los dos últimos apartados, los defensores de la PSC consideran que sus presuntas fallas (entre ellas, las explicativas) no alcanzan para asegurar la eliminación futura del marco intencional que la misma proporciona, en el caso de confrontación con una neurociencia desarrollada.⁵³ En este capítulo he mostrado, por lo tanto, que los análisis a favor y en contra de la existencia de la PSC no hacen un uso tácito ni explícito de algún principio de exclusión explicativa, ya sea el propuesto por Kim u otro, si bien las consideraciones explicativas no están ausentes en tales análisis. Quisiera hacer ahora dos consideraciones finales con respecto al alcance de estas observaciones y al lugar que creo que puede ocupar el principio de exclusión en el debate referente a la PSC.

En primer lugar, deseo destacar que el hecho de que en diversos análisis relativos a la PSC no se haga uso del principio de exclusión y se recurra a criterios epistemológicos de evaluación mas amplios, obviamente no implica una refutación de aquel. No obstante, sugiere que existe una tendencia favorable a adoptar marcos de análisis mas amplios como criterios de evaluación de teorías. Estas observaciones son consistentes con las expuestas en el capítulo III, referidas al rol que el principio de exclusión puede desempeñar como criterio de elección de teorías rivales. He señalado en ese capítulo que una objeción que puede oponerse al principio (EET), que para

Kim opera en casos como el de la PSC, consiste en afirmar que éste puede resultar admisible sólo en el caso de que las teorías explicativas en competencia resulten comparables en cuanto a los restantes criterios de aceptabilidad con respecto a los cuales pueden evaluarse. He observado, además, que Kim no parece tener en cuenta esta posibilidad cuando se refiere al caso principal de aplicación del principio, esto es, al de la PSC; tampoco, si es que considera a tal posibilidad implausible, argumenta en favor de esta tesis.

Sin embargo, creo que es posible que, a pesar de todo, algún principio de exclusión resulte relevante para la evaluación de la eliminación potencial de la PSC. Si el análisis de Kim sobre la causación mental fuese correcto,⁵⁴ unido a una visión realista de la explicación, permitiría argumentar en favor de la eliminación de la PSC, ya que implicaría que las propiedades mentales (entre ellas las creencias y deseos) no tendrían poderes causales sobre la conducta. En el caso de que esto fuese así, no se necesitaría argumentar en favor del principio del mayor poder explicativo de una neurociencia o ciencia cognitiva desarrollada; las entidades postuladas por la PSC serían eliminadas como propiedades eficaces. En tanto que objeto de estudio de una ciencia especial desaparecerían, pero no como resultado de la eliminación de la teoría que las postula debido a su inferior poder explicativo, sino porque carecerían de validez como entidades capaces de producir modificaciones en el mundo.⁵⁵ Y, al carecer la capacidad de

⁵³ Existen, por supuesto, otros análisis y argumentos tanto en favor como en contra de la existencia independiente de la PSC, pero creo que lo expuesto es suficiente para ilustrar mi argumento.

⁵⁴ *Supra*, p. 20.

⁵⁵ Podría pensarse que su eliminación se produciría debido a su supuesta irrealdad, de acuerdo con los resultados expuestos en el capítulo I. No obstante, creo que no sería necesario basarse en este supuesto para descartar a las propiedades mentales como objeto de una ciencia autónoma. En primer lugar, porque existen argumentos en contra de esta supuesta irrealdad de estas propiedades (*supra*, p. 27n). En segundo término, porque el reconocimiento de su incapacidad de ser parte activa en relaciones de determinación resulta suficiente para dejar de considerarlas entidades con poderes explicativos.

ser parte activa en alguna relación de determinación, se verían privadas de todo poder explicativo.

CONCLUSIONES

En el curso de este trabajo he aceptado la idea de que el principio de exclusión explicativa constituye una restricción general aceptable para las explicaciones. Si se admite un punto de vista realista con respecto a la explicación en general y la explicación científica en particular, parece plausible la idea de que, sea cual fuere la relación de determinación que se acepte (causal, de superveniencia, o cual fuese), existe sólo un determinante para el hecho que se quiere explicar y, por lo tanto, sólo una explicación apropiada. Al respecto, he defendido al planteo de la exclusión de una crítica que señala que el mismo conduce a una supuesta inconsistencia.

No obstante, he considerado que el planteo de Kim falla con respecto a una serie de aspectos, a saber:

- 1.No todas las teorías de la explicación son compatibles con la aceptación irrestricta del principio de exclusión. Desde el modelo nomológico-deductivo (considerado por Kim como "internalista explicativo"), pueden admitirse dos explicaciones completas e independientes de un mismo *explanandum* en los casos de sobredeterminación. Si se acepta que algunas otras teorías de la explicación califican en la categoría de internalistas, esta observación posiblemente debería hacerse extensiva a ellas.
- 2.En su formulación explicativa general constituye una extensión injustificada de su formulación explicativa causal. Es necesario establecer, para defender la primera, de que manera los diferentes tipos de determinación o dependencia se excluyen mutuamente en la producción de sucesos.
- 3.La aceptación de la versión epistemológica del principio resulta necesaria sólo si se acepta además la versión metafísica. Esto es, considero que existen argumentos

para sostener que las consideraciones de simplicidad y unificación no son siempre suficientes como para admitir la versión epistemológica del principio.

4. Los alcances epistemológicos que Kim asigna al principio no están justificados ni desde un punto de vista teórico ni desde el análisis que efectúa acerca de los casos concretos en los que este principio parece aplicarse. Para el caso de la PSC, creo que debe probarse, por medio de una argumentación del tipo de la enunciada para las propiedades mentales en general, que su ontología es radicalmente errónea, para justificar su eliminación. Un principio general de exclusión explicativa no es suficiente para justificar una movida tan drástica.

No obstante, hay ciertos aspectos del análisis de Kim que permanecen en pie: el núcleo del estudio de Kim, esto es, la exclusión causal-explicativa de las propiedades mentales, no se ha visto afectado por la crítica expuesta. Creo que es en este aspecto donde el análisis de Kim es más sólido; el resultado de la ineficacia causal de lo mental y su consecuente irrelevancia explicativa no parece poder evitarse. Sin embargo, y como línea de investigación posterior, cabe preguntarse en que medida una teoría pluralista de la explicación (es decir, una que acepte explicaciones no causales) puede evitar ese resultado indeseable para el materialismo no reductivo, y en general para las denominadas "ciencias especiales". Un análisis de esta naturaleza permitiría además, creo, ofrecer una perspectiva más clara acerca de la eliminación potencial del marco explicativo de la PSC.

Por último, existe una línea de investigación de la cual no me he ocupado y que también puede ser objeto de indagación posterior, ya que podría constituir una alternativa al eliminativismo y al no reduccionismo. Esta es la referida a los problemas que suscita la aceptación de la identidad mente-cerebro, y a la posibilidad de reducciones específicas por especie, tal como Kim la

plantea en su (1989b) y (1992). La tesis de la realizabilidad múltiple, o la del anomalismo de lo mental no protegen, según Kim, al materialismo no reductivo de la posibilidad de "reducciones locales" de la psicología a la neurofisiología. De acuerdo con lo expuesto, los partidarios del materialismo no reductivo sólo tienen abiertas tres posibilidades: o un retorno al dualismo sustancial (si desean mantener la creencia en que las propiedades mentales tienen los poderes que les atribuyen), alguna variante del reduccionismo, o el eliminativismo liso y llano. El desafío para los no reduccionistas consiste nuevamente, por lo tanto, en hallar una forma de materialismo que preserve un rol activo para las propiedades mentales sin caer en las extravagancias ontológicas (y empíricamente inadmisibles) involucradas en el dualismo de sustancias.

REFERENCIAS BIBLIOGRÁFICAS

- Bechtel, William (1988) *Filosofía de la Mente. Una panorámica para la ciencia cognitiva (Philosophy of Mind. An Overview for Cognitive Science)*, Madrid, Tecnos, 1991.
- Block, Ned (1990), "Can the Mind Change the World?", en Cynthia y Graham Macdonald (eds.), 1995.
- Bunzl, Martin (1979), "Causal overdetermination", *Journal of Philosophy* 76, 134-50.
- Churchland, Paul (1981), "El materialismo eliminativo y las actitudes proposicionales", en Rabossi, E. (comp.), 1995 ("Eliminative Materialism and Propositional Attitudes", *Journal of Philosophy* 78).
- Dretske, Fred (1990), "Does meaning matter?", en Cynthia y Graham Macdonald (eds.), 1995.
- Engel, Pascal (comp.) (1988), *Psicología ordinaria y ciencias cognitivas*, Barcelona, Gedisa, 1993.
- 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.
- Goldman, Alvin (1969), "The Compatibility of Mechanism and Purpose", *Philosophical Review* 78.
- Haack, Susan, (1995), *Evidence and Inquiry. Toward Reconstruction in Epistemology*, Oxford, Blackwell.

- Hempel, Carl (1965), "Aspectos de la explicación científica", en Carl Hempel, *La explicación científica*, Barcelona, Paidós, 1988.
- Horgan, Terence, y James Woodward (1985), "Folk Psychology is here to stay", *Philosophical Review*, 197-226.
- Jackson, Frank. y Philip Pettit (1990a), "Program explanation: A General Perspective", *Analysis*, 50.
 ----- (1990b), "In Defense of Folk Psychology", *Philosophical Studies* 59.
- 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 Kim (1993b).
 ----- (1978), "Supervenience and Nomological Incommensurables", *American Philosophical Quarterly*, v. 15.
 ----- (1979), "Causality, Identity, and Supervenience in the Mind-Body Problem", *Midwest Studies in Philosophy*, vol. 4.
 ----- (1984), "Concepts of supervenience", *Philosophy and Phenomenological Research* 45, pp. 153-176. Reimpreso en Kim (1993b).
 ----- (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.
 ----- (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 Kim (1993b).

- (1989b), "The Myth of Nonreductive Materialism",
*Proceedings and Addresses of the American
Philosophical Association* 63 (1989), pp.31-47.
Reimpreso en Kim (1993b)
- (1990), "Explanatory Exclusion and the Problem of
Mental Causation", en Cynthia y Graham Macdonald
(eds.), 1995.
- (1991) "Dretske on How Reasons Explains Behavior",
en Brian McLaughlin (ed.), *Dretske and His Critics*,
Oxford, Basil Blackwell, 1991. Reimpreso en (1993b).
- (1992) "Multiple Realization and the Metaphysics
of Reduction", *Philosophy and Fenomenological
Research* 52, pp. 1-26. Reimpreso en (1993b).
- (1993a), "The Nonreductivist's Troubles with
Mental Causation", en John Heil y Alfred Mele (eds.),
Mental Causation, Oxford, Oxford University Press,
1989. Reimpreso en Kim (1993b).
- (1993b), *Supervenience and Mind*, Cambridge,
Cambridge University Press.
- (1994a), "Second Order Properties and Mental
Causation", s/l.
- (1994b) "Explanatory Knowledge and Metaphysical
Dependence", en E. Villanueva (ed.), *Philosophical
Issues*, 5, Truth and Rationality, Atascadero, Ca.,
Ridgeview Publ. Co., 1994.
- Kuhn, Thomas (1962), *La estructura de las revoluciones
científicas*, Buenos Aires, Fondo de Cultura
Económica, 1990.
- Loeb, Louis (1974), "Causal Theories and Causal
Overdetermination", *Journal of Philosophy* 71, pp.
525-44.