

prender la radicalización preteórica de la fenomenología si no se repara en el profundo interés que Heidegger tenía por la mística en esos años.

Recibido: 04-2012; aceptado: 08-2012

UN ANÁLISIS CRÍTICO DE LA CONCEPCIÓN MECANICISTA DE LA EXPLICACIÓN

Sergio Daniel Barberis

Universidad de Buenos Aires

Consejo Nacional de Investigaciones Científicas y Técnicas

RESUMEN: En este trabajo me propongo desarrollar un estudio crítico de la concepción mecanicista de la explicación científica. En primer lugar, argumento que la caracterización mecanicista de los modelos fenoménicos (no explicativos) es inadecuada, pues no ofrece un análisis aceptable de los conceptos de *modelo científico* y *similitud*, que son fundamentales para la propuesta. En segundo lugar, sostengo que la caracterización de los modelos mecanicistas (explicativos) es igualmente inadecuada, pues los análisis disponibles de la relación explicativa de *relevancia constitutiva* implican una tesis metafísica que es rechazada por los mismos mecanicistas. Concluyo que el mecanicismo no ofrece todavía una elucidación aceptable de la explicación científica.

PALABRAS CLAVE: explicación científica, mecanicismo, modelos fenoménicos, relevancia constitutiva, causación internivel.

ABSTRACT: In this paper, I offer a critical assessment of the mechanist approach to scientific explanation. Firstly, I argue that the mechanist characterization of (non explanatory) phenomenological models is inadequate, since it does not develop an explication of the concepts of *scientific model* and *similarity*, which are indispensable to the approach. Secondly, I claim that the mechanist conception of (explanatory) mechanist models is inadequate as well, since all the available analyses of the explanatory relation of *constitutive relevance* imply a metaphysical thesis that is rejected by the mechanists themselves. I conclude that mechanicism needs to be emended if it aims to be considered as a genuinely illuminating approach to scientific explanation.

KEYWORDS: scientific explanation, mechanism, phenomenological models, constitutive relevance, interlevel causation.

1. Introducción

En este artículo me ocupo de la siguiente cuestión filosófica: ¿cuál es la naturaleza o estructura de las explicaciones en ciencias especiales? Desde hace al menos una década, existe un creciente acuerdo en la comunidad filosófica acerca de la adecuación y la fertilidad de la “concepción mecanicista de la explicación”, según la cual explicar un fenómeno consiste en exhibir las partes, las actividades y los rasgos organizacionales relevantes del mecanismo en el cual ese fenómeno está realizado. Aunque existen notables antecedentes, esta concepción tiene su origen en el artículo “Thinking about mechanisms” (2000) de Paul Machamer, Lindley Darden y Carl Craver, y encuentra su articulación conceptual más lograda en la *opus magnum* de Craver, *Explaining the Brain* (2007). Existen análisis mecanicistas de la explicación en biología celular, neurobiología, neurociencias, psicología cognitiva y en la teoría de la evolución, por sólo mencionar algunos campos.¹ En medio de esta “manía por los mecanismos” (Weiskopf 2011, p. 313), en este artículo me propongo discutir algunos de los principales problemas conceptuales que afectan a la concepción mecanicista de la explicación científica.

La estructura de este trabajo es la siguiente. Comenzaré por presentar las condiciones mínimas de adecuación que una propuesta filosófica acerca de la explicación científica debe satisfacer (sección 2). De especial importancia para este estudio resultará la condición de elucidación o aclaración conceptual. Luego expondré la manera en la cual los mecanicistas distinguen entre modelos fenoménicos y modelos explicativos, y argumentaré que la caracterización mecanicista de los modelos fenoménicos no satisface el requisito de elucidación conceptual (sección 3). A continuación,

1. Véase Bechtel y Richardson 1993; Machamer, Darden y Craver 2000; Glennan 2005; Darden 2006; Craver 2006, 2007; Bechtel 2008; Barros 2008; Kaplan 2011 y Piccinini y Craver 2011, entre otros.

dedicaré la mayor parte del artículo a reconstruir y evaluar críticamente la concepción mecanicista de los modelos explicativos (sección 4), centrándome en la caracterización de tres aspectos fundamentales: el fenómeno *explanandum*, las actividades causalmente relevantes y las partes constitutivamente relevantes de un mecanismo. Argumentaré que la elucidación mecanicista de las relaciones de relevancia constitutiva (entre un mecanismo como totalidad y sus partes componentes) no satisface la condición de elucidación conceptual. Por un lado, los mecanicistas rechazan la posibilidad de la causación entre ítems pertenecientes a distintos niveles de un mecanismo. Pero, por otro lado, los análisis que ofrecen de la relevancia constitutiva, ya sea en términos de mutua manipulabilidad o de condiciones necesarias y/o suficientes, implican que las relaciones constitutivas entre ítems de distintos niveles son relaciones causales. Concluiré que, a pesar del entusiasmo por el mecanicismo en la filosofía actual de las ciencias especiales², esta concepción de la explicación nos debe todavía una respuesta a estos serios problemas conceptuales.

2. El mecanicismo y el requisito de elucidación conceptual

Siguiendo el esquema general propuesto por José Díez (2008, p. 385), sostengo que un análisis de la explicación debe ofrecer, al menos, una especificación de: (a) en qué consiste el *explanandum* (aquello que debe ser explicado), (b) en qué consiste el *explanans* (aquello que proporciona la explicación) y (c) cuál es la relación explicativa, esto es, cuál es la relación entre el *explanans* y el *expla-*

2. Especificar en qué consisten las llamadas “ciencias especiales” no es una tarea trivial (cf. Fodor 1974). La idea intuitiva es que existe una distinción entre aquellas ciencias fundamentales que estudian los objetos, procesos y leyes más básicos –i. e., la física de partículas– y aquellas otras ciencias, las especiales, que estudian tipos específicos de fenómenos pertenecientes a niveles superiores (menos básicos), tales como los fenómenos químicos, biológicos o psicológicos. La filosofía de las ciencias especiales se ocupa de los aspectos metodológicos, epistemológicos, metafísicos, etc. específicos de estas últimas disciplinas.

nandum que hace verdadera la proposición según la cual lo primero explica lo segundo. Una propuesta que siga este patrón de análisis debe ser evaluada, mínimamente, de acuerdo con los siguientes requisitos o criterios:

(1) *Adecuación extensional*: el análisis filosófico propuesto no debe excluir casos paradigmáticos de explicación ni debe incluir casos claros de no explicación en ciencias especiales.

Este requisito es introducido en estos términos por Díez (2008, p. 396) y es aceptado explícitamente por Craver (2007, p. 19) cuando señala que el análisis de la explicación debe ser “descriptivamente adecuado”. Sin embargo, debemos atender a la siguiente aclaración. En varios lugares de la literatura mecanicista se menciona, como un “argumento” a favor de dicha concepción, el hecho de que, en la mayoría de las explicaciones que aparecen en artículos de revistas biológicas especializadas, no se mencionan leyes de la naturaleza ni aparecen descripciones de fenómenos a partir de leyes científicas (tal como podría esperar algún intérprete incauto del modelo inferencialista nomológico de Hempel), sino que “al examinar la literatura biológica, queda claro que el término que los biólogos usan frecuentemente en contextos explicativos es ‘mecanismo’” (Bechtel y Abrahamsen 2005, p. 422). Quisiera señalar que no es este tipo de adecuación literal el que se busca en el requisito de adecuación extensional. Que “una búsqueda de títulos de artículos en revistas que incorporen ‘mecanismo’ y ‘síntesis de proteínas’ produzca 203 resultados”, en tanto que “una búsqueda con ‘ley’ y ‘síntesis de proteínas’ no arroje ninguno” (*ibídem*) no constituye evidencia concluyente (ni mucho menos) a favor de la concepción mecanicista. Lo que el filósofo de la ciencia se propone es ofrecer una elucidación o *reconstrucción racional* de las prácticas o pautas explicativas en el ámbito científico de su interés (Craver 2007, p. 19), que no tiene por qué recoger los términos en los cuales los científicos reflexionan (de manera más o menos intuitiva) acerca de sus propias prácticas.

(2) *Elucidación conceptual*: el análisis filosófico propuesto debe resultar aclaratorio: los conceptos primitivos que se utilicen en el análisis no deben ser más oscuros que el concepto mismo de *explicación* que se pretende elucidar.

Este segundo requisito también es introducido en estos términos por Díez (2008, p. 386) y es aceptado implícitamente por Craver (2007, p. 20) cuando establece que todo análisis de la explicación debe ofrecer una “demarcación” fundamentada entre las explicaciones genuinas y otros tipos de logro científico. Por ejemplo, un análisis adecuado debe distinguir, de una manera razonada, entre la simulación y la explicación de un fenómeno, así como entre esta última y la construcción de meras taxonomías o categorizaciones. Tal fundamentación sería imposible si los conceptos con los cuales se analiza la explicación resultasen más oscuros que el concepto de *explicación* mismo, esto es, si la propuesta no fuese esclarecedora.

A partir del modo en el cual se conciben cada uno de los componentes (a)-(c) de un análisis de la explicación, es posible construir una clasificación de las propuestas filosóficas disponibles en la literatura sobre el tema. Siguiendo parcialmente a Salmon (1984, 1998) cabe distinguir entre las concepciones epistémicas y las concepciones ónticas de la explicación. Según las primeras, los términos de una explicación (aquello que se vincula mediante una relación explicativa) son, en sí mismos, de naturaleza epistémica; pueden tratarse, por ejemplo, de proposiciones, ecuaciones matemáticas, modelos científicos o partes de modelos científicos, etc. La relación explicativa entre esos términos también es epistémica: por ejemplo, pueden ser relaciones de consecuencia lógica, apoyo inductivo, subsunción (modeloteórica), unificación, etc. Según la caracterización más general que puede hacerse de las concepciones epistémicas, éstas sostienen que una explicación científica es un conjunto finito de operaciones sobre objetos representacionales (Wright, *inédito*). El caso paradigmático de este tipo de propuesta es la concepción inferencialista nomológica de Hempel (1965), según la cual una explicación es un argumento o inferencia, los *relata* de la relación explicativa son enunciados que funcionan como premisas y conclusión (enunciados de ley y/o condiciones iniciales en el *explanans*; descripciones de hechos o leyes en el *explanandum*) y la relación explicativa es una relación de previsibilidad nómica. Existen otras propuestas que satisfacen la caracterización general de Wright y que ofrecen análisis alternativos al de Hempel. La más conocida es el análisis de la explicación como unificación de Kitcher (1981). Otras alternativas epistémicas son, por ejemplo, la

propuesta conexionista de la explicación como procesamiento paralelo distribuido de Churchland (1989), el modelo erotético de van Fraassen (1980) y el análisis de la explicación como subsunción de Bartelborth (1996) y Díez (2002, 2008).

Según las concepciones ópticas, las explicaciones son, ante todo, aspectos objetivos del mundo, independientemente de cómo nos lo representemos (Salmon 1984, p. 274). Tanto el *explanandum* como el *explanans* son eventos, objetos y/o propiedades en el mundo (Salmon 1998, p. 54): el *explanandum* puede consistir, por ejemplo, de eventos en el mundo o de ciertas capacidades de un sistema físico; el *explanans* puede consistir de procesos causales antecedentes o de mecanismos en el mundo. La relación entre *explanans* y *explanandum* es una relación metafísica (o “natural”). Pueden tratarse, según la propuesta del caso, de relaciones de causación, realización, identidad, relevancia constitutiva, etc. El punto en común entre las concepciones ópticas es, supuestamente, el rechazo de lo que Salmon (1998, p. 95) llamó “el tercer dogma del empirismo”: la idea de que las explicaciones científicas son argumentos.

Para la concepción mecánico-causal de la explicación de Salmon, explicar un evento no consiste en inferir su descripción de premisas cualesquiera, sino en indicar o exhibir su posición en la estructura causal del mundo (Salmon 1984, 22; 1998, 66). Para explicar etiológicamente un evento *E* (siendo que *E* ocupa un volumen finito en el espacio-tiempo), debemos ostentar aquellos procesos e interacciones causalmente relevantes que ocupan el cono de luz pasado de *E* (Salmon 1984, 275). Si, en cambio, lo que queremos explicar es por qué *E* exhibe ciertas características, debemos ostentar aquellos procesos e interacciones causales relevantes del mecanismo interno (esto es, ubicado dentro de los límites del volumen espaciotemporal de *E*) que constituye la naturaleza del evento en cuestión. A éste último tipo de pauta explicativa Salmon la denomina “explicación constitutiva” (1984, 270). El principal desafío para la concepción mecánico-causal radica en la satisfacción del requisito de elucidación conceptual: debe proporcionarnos un análisis de las nociones de *causación* y *relevancia causal* que resulte tanto o más esclarecedor que la noción de *explicación científica*. La respuesta de Salmon a este desafío está desarrollada en la teoría de la causación como transmisión de una cantidad conservada (Salmon 1998). Según esta teoría, un proceso es

causal si transmite una cantidad que se conserva en el curso de una interacción, esto es, que tiene el mismo valor antes y después de dicha interacción (por ejemplo: el momento lineal o la carga eléctrica). Una interacción es causal si involucra el intercambio de cantidades conservadas.³

Pues bien, la concepción mecanicista de la explicación puede verse como una ampliación y una corrección del modelo mecánico-causal de Salmon, manteniendo la intuición básica según la cual explicar un fenómeno es exhibir su lugar en la estructura causal del mundo (Craver 2007). En primer lugar, se trata de una “ampliación” pues los mecanicistas pretenden incorporar como legítimas (desde una perspectiva óptica de la explicación) a las pautas explicativas de las ciencias especiales, con especial interés en las explicaciones de las ciencias biológicas y de las ciencias cognitivas. El concepto de mecanismo de Craver, Glennan y otros filósofos mecanicistas se propone dar cuenta de la explicación de fenómenos complejos, como el funcionamiento de las células o del cerebro –fenómenos que los análisis tradicionales de la explicación, incluido el de Salmon, ni siquiera tomaron en cuenta. En segundo lugar y dado este primer objetivo, la atención de los mecanicistas está puesta, no tanto en las explicaciones etiológicas de un fenómeno, como en las mencionadas explicaciones constitutivas o componentes. En este sentido, la relación explicativa fundamental no será la de interacción causal, como en el análisis de Salmon, sino la noción de relevancia constitutiva. En tercer lugar, el mecanicismo representa una “corrección” de la propuesta de Salmon porque, si bien esta última especifica ciertos criterios para determinar si un proceso pertenece a la causa total de un evento *E*, no discute explícitamente la cuestión de qué condiciones debe satisfacer un proceso perteneciente a la causa total para contar, además, como un antecedente *relevante* en el contexto de una demanda de explicación particular. Los mecanicistas se proponen, entonces, “corregir” el modelo de Salmon pues ofrecen una elucidación de la noción de relevancia explicativa que no depende necesariamente del intercambio de cantidades fundamentales y que es aplicable, en prin-

3. Sobre este tema, véase “Las actividades causalmente relevantes” de la sección 4 de este artículo.

cipio, a las relaciones constitutivas en los mecanismos postulados por las ciencias especiales.

Con cierta frecuencia, los mecanicistas compendian las conocidas falencias del análisis de Hempel y las utilizan como evidencia *prima facie* a favor de la concepción óptica.⁴ Sin embargo, es importante notar que varias concepciones epistémicas rechazan también el “tercer dogma”, sin por ello adoptar la posición óptica (cf. Díez 2002, 2008). En este punto, la manera en la cual los mecanicistas usualmente reconstruyen el escenario filosófico es, cuanto menos, confusa. Lo que caracteriza a las concepciones ópticas es, más bien, la desestimación de los factores epistémicos como factores imprescindibles o prioritarios en la caracterización de la relación explicativa y de los términos que dicha relación vincula. Así, Craver se pregunta:

¿Qué es lo que explica la liberación de neurotransmisores? El potencial de acción, el influjo de iones de calcio, [etc.]. Existen mecanismos (las explicaciones objetivas) y existen sus descripciones (los textos explicativos). Las explicaciones objetivas no son textos; son objetos concretos. Son hechos, no representaciones. (Craver 2007, p. 27)

Esta distinción entre “explicaciones objetivas” y “textos explicativos” permite resolver, en principio, una notoria tensión presente en la literatura mecanicista. Por un lado, los mecanicistas se proponen ofrecer un análisis metafísico (al menos, no epistémico) de la explicación; por otro lado, en muchas de las presentaciones canónicas de esta concepción, la discusión gira en torno de las condiciones que deben cumplirse para que un determinado *modelo científico* sea considerado mecanicista y, por lo tanto, explicativo.⁵ Así, Craver afirma que:

Los modelos desempeñan muchos papeles en ciencia más allá de proveer explicaciones. Se utilizan para hacer predicciones precisas y exactas. Se utilizan para resumir datos. Se usan como heurística en el diseño de experimentos. Se utilizan para demostrar consecuencias sorpren-

4. Véase especialmente Bechtel y Abrahamsen 2005, Bechtel y Wright 2007 y Bechtel 2008.

5. Véase Woodward 2002, Craver 2007 y Kaplan 2011.

denes y contraintuitivas de formas particulares de organización sistemática. Pero *algunos modelos tienen una propiedad adicional: son explicaciones*. (Craver 2007, p. 355; mis cursivas)

Los modelos científicos pueden llegar a ser explicativos, y ellos son los elementos epistémicos por antonomasia. ¿Cómo conciliar estas afirmaciones con el rechazo de las concepciones epistémicas de la explicación? Una manera consiste en sostener que existe una asimetría entre las explicaciones objetivas y los textos explicativos. Para Craver, las explicaciones objetivas (*qua* objetos concretos) son el *punto de partida* correcto del análisis filosófico. El mecanicismo está centrado en el fenómeno del mundo que se pretende explicar, en los componentes del mecanismo en el mundo que realiza dicho fenómeno, y en la relación metafísica que existe entre ambos. El estudio de las propiedades explicativas de los modelos científicos es subsidiario o dependiente del análisis metafísico de las relaciones de constitución entre los componentes de un mecanismo y los fenómenos. De hecho, a pesar de que “quizá existan muchas cosas interesantes que decir” acerca de los modelos científicos (Craver 2007, p. 27n), lo único que le interesa destacar al mecanicista es que el carácter explicativo, o no, de un modelo científico depende crucialmente de que ese modelo logre *representar* adecuadamente las partes reales, sus actividades y la organización del mecanismo en el mundo. Dicho crudamente, el estudio de las propiedades “internas” o “intrínsecas” de los modelos resulta ocioso para el análisis de la explicación, pues la única propiedad de los modelos que es determinante para dicha cuestión es la propiedad relacional de representar adecuadamente los objetos concretos, primariamente explicativos, en el mundo (cf. Matthewson y Calcott 2011).

3. Modelo fenoménicos

Hemos visto que, para Craver, existen algunos *modelos* científicos que, aunque útiles para ciertos aspectos de la práctica científica, no son explicativos, mientras que otros modelos, en la medida en que cumplan ciertas condiciones representacionales específicas, pueden llegar a ser genuinamente explicativos. Lo cierto es que no existe una elucidación ampliamente aceptada del concepto de *mode-*

lo científico, no sólo al interior del mecanicismo, sino en la filosofía de la ciencia en general. Los principales exponentes del mecanicismo no se comprometen con ningún análisis en particular y evitan tomar partido en los debates acerca de la naturaleza de los modelos científicos (¿son entidades abstractas? ¿son entidades concretas no actualizadas?, cf. Godfrey-Smith 2006, Matthewson y Calcott 2011), o acerca del carácter más o menos autónomo de los modelos respecto de las teorías y respecto de la evidencia empírica (cf. Morgan y Morrison 1999), o respecto de la relación que debe vincular a los modelos con los sistemas que pretenden modelar (cf. Giere 1999; Suárez 2003).⁶ Lo que queda claro, al menos en la obra de Craver, es que no está usando el concepto matemático de “modelo” de la teoría de modelos, tal como se lo utiliza en lógica, matemática y en la concepción estructuralista de las teorías. Craver (2002) está interesado en rescatar, justamente mediante su idea de *modelo mecanicista*, aquellos “patrones no formales” de la teorización científica que, supuestamente, no son tenidos en cuenta por los filósofos interesados en la reconstrucción formal de las teorías científicas. De manera controversial, los mecanicistas parecen considerar que el concepto de *modelo científico* se halla lo suficientemente comprendido como para ser tomado como un primitivo en el análisis de la explicación.

De acuerdo con la taxonomía general de modelos científicos propuesta en Craver (2006, 2007) cabe distinguir, en primer lugar, entre los *modelos representacionales* y los *modelos ficcionales* o “modelos como si”. Los primeros pretenden ser descripciones más o menos abstractas de sistemas reales. Los segundos, en cambio, son meras “ficciones útiles” que no tienen ninguna pretensión representacional. Si bien Craver no dice nada acerca de los criterios para realizar la distinción, podría pensarse que trata de emular la distinción clásica entre “modelos implícitos” y “modelos analógicos” (Nagel 1961). Mientras los modelos implícitos (o representacionales) pretenden describir, de alguna manera, los fenómenos a los cuales se aplican, los modelos analógicos (o ficcionales) tienen una finalidad puramente heurística o pedagógica y no acarrear ningún compromiso ontológico respecto de las entidades y procesos que allí aparecen. Un ejemplo de modelo científico “como si” es la

6. Por ejemplo, Bechtel 2008, Craver 2007 y Glennan 2005.

representación de cierto fenómeno eléctrico como un sistema hidrodinámico: si bien puede resultar útil para comprender algunas propiedades de una corriente eléctrica, nadie piensa que literalmente un sistema eléctrico es un sistema hidrodinámico.

Dentro de los modelos representacionales Craver distingue, ahora sí, entre los modelos (meramente) fenoménicamente adecuados –o modelos fenoménicos *tout court*– y los modelos genuinamente explicativos. Si el mapeo *input-output* en un modelo *S* resulta suficientemente similar (según ciertos criterios contextuales de relevancia) al mapeo *input-output* en un sistema *T* en el mundo, entonces puede afirmarse que *S* es fenoménicamente adecuado, o que *S* es un modelo fenoménico de *T* (Craver 2006). Sintetizando los aspectos fundamentales de las propuestas de varios filósofos mecanicistas⁷, considero que resulta adecuado reconstruir la caracterización mecanicista de los modelos fenoménicos de la siguiente manera:

[FEN]

Un modelo científico *F* de un fenómeno φ es un modelo fenoménico para φ si y sólo si: (1) las variables de *input* en *F* representan las condiciones de *input* en φ ; (2) las variables de *output* en *F* representan las condiciones de *output* en φ ; (3) las dependencias postuladas entre las variables de *input* y *output* de *F* son similares, de manera relevante, a las dependencias causales entre las condiciones de *input* y *output* en φ .

Los modelos fenoménicos, aunque importantes para la predicción y el compendio de regularidades empíricas –entre otras prácticas epistémicas, cf. Bogen (2005)– no son explicativos. Los modelos explicativos genuinos son aquellos que describen, de manera más o menos precisa, el mecanismo responsable de las diversas regularidades que se dan en el nivel del fenómeno. Algunos de los ejemplos de modelos fenoménicos mencionados por los mecanicistas son: el modelo de Ptolomeo de los movimientos aparentes de los planetas, la ley de Snell, la fórmula de Balmer, los modelos computacionales oracionales de la mente, el modelo conexionista de la

7. En particular, me baso en las propuestas de Craver (2006, 2007), Glennan (2005) y Kaplan (2011).

generación de verbos en pasado de Rumelhart y McClelland (1986) y, de manera polémica, el modelo electrofisiológico del potencial de acción de Hodgkin y Huxley (1952).

Según se desprende de FEN, entonces, los modelos fenoménicos son modelos científicos que, aunque “salvan los fenómenos”, no describen las partes componentes de ningún mecanismo y, por lo tanto, no son explicativos. Está elucidación, argumentaré, no está exenta de problemas conceptuales.

En primer lugar, no queda claro de qué manera los modelos representacionales fenoménicos deben distinguirse de los modelos no representacionales o ficcionales. Un modelo conexionista de la producción de verbos (uno que cuente con suficiente confirmación empírica, en particular), postula toda una serie de estratos de redes de nodos, asigna pesos a cada nodo y especifica una serie de ecuaciones que rigen los cambios en los pesos asignados a los nodos de la red de acuerdo con los cambios en el *input* (Rumelhart y McClelland 1986). Si bien los conexionistas consideran que los modelos de este tipo son biológicamente más plausibles que los modelos oracionales de la mente, nadie estaría dispuesto a afirmar que los nodos de la red representan directamente partes reales del mecanismo relevante en el cerebro. ¿Cuál es la diferencia sustantiva, entonces, entre afirmar que un modelo conexionista es fenoménicamente adecuado aunque no describe partes reales del mecanismo y afirmar que se trata, meramente, de un modelo “como si”, de una ficción útil a los fines de la predicción o de la sistematización de la evidencia empírica? Parece que los únicos modelos genuinamente representacionales son los modelos mecanicistas, pues son los únicos modelos que pueden tener relaciones representacionales exitosas con el mundo (“describen las partes reales del mecanismo”). En el caso de los modelos fenoménicos, puesto que las “partes componentes” que postulan no tienen contraparte óptica en el mecanismo, no parece que pueda hablarse de una relación representacional exitosa y, por lo tanto, no hay razón por la cual evitar hablar de meras “ficciones útiles”.

Considérese ahora la condición (FEN-3) del análisis mecanicista, según la cual los modelos representacionales fenoménicos son aquellos que establecen un mapeo *input-output* lo suficientemente similar al mapeo *input-output* del fenómeno en cuestión. ¿Puede el concepto de *similitud* ofrecernos un análisis satisfactorio de la idea

de que un modelo fenoménico representa el fenómeno empírico aunque no representa los mecanismos subyacentes?

Mi tesis es que estamos frente a un caso en el cual se elucida lo oscuro por lo más oscuro. Existen varios argumentos, bien conocidos, en contra de los análisis de la relación de representación científica en términos de similitud. Muchos de ellos aparecen reseñados en Suárez (2003) y Frigg (2006). Me interesa particularmente el llamado “argumento lógico”, que se remonta a la crítica de Goodman (1976) a las teorías de la similitud y que es retomado por Suárez. El fenómeno de la representación en general (y, en particular, en el caso de las representaciones científicas) es usualmente concebido como una relación no reflexiva y no simétrica⁸. Un conjunto de ecuaciones en un modelo biológico puede representar la evolución de un rasgo en una población a través de varias generaciones, pero ese conjunto de ecuaciones no tiene por qué representarse a sí mismo (no reflexividad) ni la evolución de la población representa las ecuaciones (no simetría). Sin embargo, las relaciones de similitud son reflexivas y simétricas. Un hombre, por ejemplo, es similar a sí mismo, y es similar a cualquier otro hombre que sea similar a él (Suárez 2003, p. 233). Reemplazar la relación de similitud por una relación de *isomorfismo* no mejora la situación: basta con señalar que se trata de una relación de *equivalencia*, i. e., reflexiva, simétrica y transitiva.

Ni los mecanicistas, ni los filósofos de la ciencia basada en modelos parecen interesados en responder estos argumentos.⁹ Incluso autores como Giere (1999) y Godfrey-Smith (2006, p. 733) se niegan explícitamente a ofrecer un análisis formal de las relaciones de similitud y se limitan a apelar a una noción informal de similitud en el tratamiento filosófico de las relaciones entre un modelo y un sistema en el mundo. Lo cierto es que nadie les exige un análisis formal de la relación de similitud; lo que el argumento lógico exige es que, aún en términos informales, indiquen de qué manera la propuesta permite explicar ciertas propiedades ampliamente

8. Plausiblemente, la relación de representación sea además irreflexiva, asimétrica e intransitiva, pero no se requiere argumentar a favor de dicha caracterización a los fines del presente argumento (cf. Suárez 2003, p. 232).

9. Véase Giere 1999, Godfrey-Smith 2006 y Matthewson y Calcott 2011.

aceptadas de la noción de representación científica. El mecanicismo no ofrece un análisis aceptable del concepto de similitud, y ello le impide satisfacer el requisito de elucidación conceptual en su caracterización de los modelos fenoménicos.

4. Modelos mecanicistas

Los modelos fenoménicos deben poder distinguirse de los modelos genuinamente explicativos. Craver (2006, p. 358), y luego otros (Kaplan 2011, p. 354), ofrecen, en primer término y de manera provisoria, una “defensa instrumentalista” de la distinción entre esos dos tipos de modelos. Según esta defensa, los modelos explicativos son más útiles que los fenoménicos a los fines del control y la manipulación de los fenómenos. Esta idea aparece originalmente en Woodward (2003), como base para la distinción entre explicación y descripción:

Tenemos los comienzos, al menos, de una explicación cuando hemos identificado factores o condiciones tales que las manipulaciones o los cambios sobre dichos factores o condiciones producen cambios en el resultado que se pretende explicar. El conocimiento descriptivo, por contraste, es conocimiento que, a pesar de que puede proveer una base para la predicción, la clasificación o la representación, o la sistematización más o menos unificada, no provee información que sea potencialmente relevante para la manipulación. En esto radica el contraste principal entre la explicación causal y la descripción (Woodward 2003, p. 10)

Según estos autores, los modelos explicativos son más útiles para la manipulación y el control porque permiten responder a un rango más amplio de preguntas contrafácticas del tipo: ¿qué pasaría si las cosas fuesen diferentes? O *preguntas-w*. Por supuesto, los modelos fenoménicos también permiten responder un rango acotado de preguntas-w, especialmente aquellas vinculadas con posibles modificaciones en las condiciones de *input* del sistema. Por ejemplo, el modelo de Hodgkin y Huxley (1952) del potencial de acción permite responder preguntas-w del tipo: “¿Qué pasaría con el potencial de la membrana si la concentración de potasio extracelular tuviese tal o cual valor?” (Kaplan 2011, p. 354). El punto es

que los modelos explicativos permiten responder un rango mucho más amplio de preguntas contrafácticas, aquellas que involucran intervenciones en las condiciones de *input* y, de manera crucial, intervenciones en las diversas partes componentes del mecanismo representado por el modelo. El acento no debería ponerse, entonces, sobre el rango o la cantidad de intervenciones que uno y otro tipo de modelos posibilitan, sino sobre la variedad de las intervenciones que cada uno ofrece. Se trata de una diferencia cualitativa entre los modelos fenoménicos y los explicativos justamente porque estos últimos permiten responder un tipo de preguntas-w que está vedado a los primeros: aquellas preguntas-w relativas a las modificaciones en el comportamiento del fenómeno a partir de intervenciones sobre las partes, las actividades y la organización del mecanismo subyacente.

Llegamos así al criterio “definitivo” para distinguir los modelos explicativos de los fenoménicos: *los modelos explicativos son aquellos que representan mecanismos* (Craver 2006, p. 367), esto es: “describen cómo las entidades y las actividades constituyentes están organizadas para exhibir un fenómeno” (Craver 2007, p. 122). Es porque dichos modelos describen mecanismos que el criterio provisoria, en términos de manipulación y control, puede funcionar. Sintetizando las propuestas de análisis de los principales representantes de la corriente¹⁰, podemos caracterizar los modelos mecanicistas de la siguiente manera:

[MEC]

Un modelo *M* de un fenómeno φ es un *modelo explicativo* de ese fenómeno si y sólo si: (1) *M* es fenoménicamente adecuado (es un modelo fenoménico de φ); (2) las “variables de partes componentes” en *M* representan las partes reales que son constitutivamente relevantes del mecanismo que realiza φ ; (3) las “actividades” o dependencias entre las variables de partes en *M* representan las dependencias causalmente relevantes entre las partes componentes del mecanismo que realiza φ ; (4) los rasgos organizacionales entre las variables de partes y actividades en *M* representan la organización de partes y actividades en el mecanismo que realiza φ .

10. Nuevamente, me apoyo en las propuestas de Craver (2007), Glennan (2005) y Kaplan (2011).

Cada uno de los componentes de esta caracterización debe, a su vez, explicitarse, y ponderarse detalladamente en relación con el requisito de elucidación conceptual.

El fenómeno *explanandum*

En primer lugar, me ocuparé de aquello que debe ser explicado mediante un modelo mecanicista: el llamado “fenómeno *explanandum*” φ . Según Glennan (2002, 2005), los límites de un mecanismo (es decir, qué porciones de la estructura causal del mundo serán consideradas como formando parte del mecanismo) están fijados por referencia al fenómeno que ese mecanismo explica. Ésta es la llamada “ley de Glennan”: si el objetivo es proveer una explicación mecanicista, el fenómeno debería estar delimitado de tal manera que corresponda con un *único* mecanismo subyacente (Craver 2007, p. 122). De esta dependencia, *prima facie* inocente, se sigue al menos una consecuencia que vale la pena consignar. Según estos autores, no puede ofrecerse un modelo mecanicista y explicativo de un fenómeno que sea, él mismo, ficticio, ni puede ofrecerse un modelo explicativo de un fenómeno que pueda ser dividido, de hecho, en otros, a cada cual correspondiéndole un mecanismo distinto. Esta consecuencia resulta altamente contraintuitiva. Consideremos un ejemplo del propio Craver. Descartes creía que los nervios eran conductos huecos que transportaban a los espíritus animales y que estos últimos activaban los músculos inflándolos. Si Descartes hubiese ofrecido un modelo del mecanismo por el cual los nervios insuflaban espíritus animales en los músculos, dicho modelo no hubiese sido un modelo mecanicista –a pesar de las apariencias y de las pretensiones mecanicistas del propio Descartes– pues el supuesto fenómeno que venía a explicar era ficticio. Recuérdese que no hay ninguna propiedad “intrínseca” de los modelos científicos que los vuelva mecanicistas. Lo único relevante es la propiedad relacional de representar un mecanismo real en el mundo. Por lo tanto, desde el mecanicismo, resulta conceptualmente imposible ofrecer (o haber ofrecido) un modelo mecanicista de un fenómeno inexistente (o que resultó inexistente según nuestro conocimiento científico actual).

Un aspecto menos polémico de la caracterización mecanicista

del fenómeno *explanandum* es el énfasis en el carácter multifacético de este último. Este énfasis viene a reparar una elucidación ampliamente citada de Robert Cummins, según la cual el fenómeno a explicar en un “análisis funcional” es una capacidad o disposición φ de un sistema o totalidad S , capacidad que puede ser descripta mediante “una ley especial que vincule las condiciones de precipitación con las manifestaciones, esto es, que especifique las condiciones de *input-output*” (Cummins 1983, p. 53). Según los mecanicistas, esta elucidación subestima la importancia que pueden tener los detalles acerca de las diferentes facetas del fenómeno a la hora de completar un mero bosquejo o esquema de un mecanismo, o de decidir entre dos modelos científicos en competencia. Así, una descripción adecuada del fenómeno φ debería incluir: (i) las condiciones estándar de precipitación y las manifestaciones estándar de φ ; (ii) las condiciones de inhibición de φ , aquellas que, de producirse, impiden la manifestación de φ ; (iii) las condiciones de modulación de φ , cuya variación altera algunas de las propiedades del fenómeno; (iv) las condiciones no estándar o artificiales de precipitación de φ (condiciones de laboratorio); (v) los efectos colaterales de φ , que, aunque no son funcionalmente significativos, pueden llegar a resultar cruciales a la hora de decidir entre modelos científicos alternativos para un mismo fenómeno.

La satisfacción de las condiciones (i)-(v) garantiza que nos encontramos frente a una descripción completa del fenómeno, y nos asegura también que la explicación mecanicista no fallará porque estemos pretendiendo explicar un fenómeno ficticio o mal individuado. Es más, Craver (2007, p. 128) llega a afirmar que: “caracterizar el fenómeno correcta y completamente es un paso crucial para convertir un análisis funcional en una explicación mecanicista aceptable”.

Ahora bien, ¿cómo sería posible satisfacer la condición (v)? Recuérdese que puede tratarse (y generalmente se trata) de efectos incidentales, no funcionales, del fenómeno. ¿Cómo podríamos describirlos completamente? Mi sospecha es que la lista de dichos efectos no puede realizarse de antemano y que, por lo tanto, no puede funcionar como “un paso crucial” para superar, de una vez, la construcción de modelos funcionales y entrar en el camino seguro de los modelos mecanicistas. Por supuesto, un mecanicista podría replicar que la idea de una descripción completa y correcta del

fenómeno es un ideal regulativo y que, en general, la comprensión de un fenómeno “coevoluciona” con nuestra comprensión de los mecanismos subyacentes (Craver 2007, 123n). Esta última afirmación es aceptable, pero entonces la descripción completa del fenómeno no es tanto un prerrequisito para desarrollar explicaciones mecanicistas aceptables, sino un punto de llegada de la explicación.

Las actividades causalmente relevantes

Para que un modelo M explique un fenómeno en sus múltiples facetas debe describir completamente, entre otras cosas, las partes componentes y las actividades del mecanismo en el cual se realiza ese fenómeno. Estas ideas están recuperadas en las condiciones (MEC-2) y (MEC-3). Considérese (MEC-3): las “actividades” o dependencias entre variables de partes componentes en M deben representar las dependencias *causalmente relevantes* entre los componentes del mecanismo en cuestión. Para que (MEC-3) resulte aceptable, en lo que concierne al requisito de elucidación conceptual, el mecanicista debe ofrecernos una caracterización satisfactoria del concepto de relevancia causal.

Podría pensarse, en primera instancia, que puede resultar productivo analizar las relaciones de relevancia causal en términos de la teoría de la causación como transmisión de Salmon (1998). De acuerdo con este autor, un proceso es causal si transmite una cantidad conservada y una interacción de dos procesos cuenta como causal si involucra el intercambio de cantidades conservadas. Machamer, Darden y Craver (2000, p. 3) no están muy lejos de adoptar esta teoría cuando afirman que las actividades que componen un mecanismo pueden ser mecánicas, electroquímicas, energéticas o electromagnéticas. Sin embargo, en la medida en que las cantidades conservadas son aquellas que se encuentran sólo en los niveles físicos fundamentales, la propuesta de Salmon resulta escasamente informativa para los casos de causación que interesan en las ciencias especiales. En primer lugar, muy pocas generalizaciones causales en biología, en neurociencias, o en psicología experimental, apelan directamente al intercambio de cantidades conservadas fundamentales (Craver 2007, p. 77). En segundo lugar, aún cuando Salmon ofrezca un análisis de la causación, no nos dice

nada acerca de qué condiciones debe cumplir un componente de la causal total de un fenómeno para ser considerado como *relevante* explicativamente (Díez 2008, p. 296). En el cono de luz pasado de un evento hay una multiplicidad de procesos e interacciones que son causas parciales del mismo, pero sólo algunos de ellos son etiológicamente relevantes en el contexto de una demanda de explicación.

En Craver (2007), el autor despliega un análisis de la causación que viene a rectificar estas falencias de la propuesta de Salmon y que permite otorgar un significado preciso al concepto de relevancia causal que aparece en la condición (MEC-3). El análisis de Craver retoma los lineamientos generales de la teoría manipulacionista de la explicación causal de Jim Woodward (2003). Según Craver:

Una variable X es causalmente relevante para la variable Y en las condiciones W si alguna *intervención ideal* sobre X en las condiciones W cambia el valor de Y (o la distribución de probabilidades sobre los posibles valores de Y). (Craver 2007, p. 94; mis cursivas)

Tanto en Woodward (2003) como en Craver (2007), los *relata* de una relación causal son caracterizados de la manera más abstracta (y por lo tanto, neutral) posible, como “variables”: determinables que son capaces de tomar valores determinados. De esta manera, los autores pretenden que el análisis de la relevancia causal sea compatible con varias caracterizaciones posibles de la causación, ya sea como una relación entre eventos, procesos, objetos, hechos o clases de contraste. Esta compatibilidad pone en evidencia que el análisis manipulacionista no ofrece una teoría (reduccionista o no) acerca de cuál es la naturaleza de la causación; sino que nos permite decidir cuándo un ítem es etiológicamente *relevante* respecto de un fenómeno. Un ítem X será causalmente relevante respecto de Y si una intervención ideal I sobre X , en el contexto adecuado, cambia el valor de Y .

El concepto fundamental en esta propuesta es el de *intervención etiológica ideal*, cuya reconstrucción definitiva es la siguiente:

[I]

I es una intervención sobre X con respecto a Y si y sólo si: (1) I no cambia Y directamente; (2) I no cambia el valor de ningún intermediario causal S entre X e Y , si no es por medio de un cambio en el valor de

X; (3) *I* no está [vinculada] con ninguna otra variable *M* que sea una causa de *Y*; (4) *I* actúa como un “interruptor” que controla el valor de *X* independientemente del valor que tome cualquier otro miembro del conjunto *U* de las causas de *X*. (Craver 2007, p. 96)

Al menos en este artículo, no estoy interesado en criticar esta elucidación en términos de intervenciones etiológicas ideales. Considero, como muchos otros filósofos de la ciencia, que constituye una de las propuestas más prometedoras de análisis de la relevancia causal. Me interesa señalar aquí que, una vez que contamos con el concepto de *intervención etiológica ideal*, podemos encarar la cuestión de qué es lo que determina que una dependencia entre algunos componentes de un mecanismo sea causalmente relevante: en el contexto de un modelo de un mecanismo, hay una actividad causalmente relevante conectando una parte componente *X* con otra parte *Y* si y sólo si es posible alterar *Y* manipulando *X*; esto es, si y sólo si es posible alterar algún rasgo de *Y* interviniendo (idealmente) sobre *X*.

Las partes constitutivamente relevantes

La condición (MEC-2) del análisis mecanicista de los modelos explicativos establece que las partes que aparecen en el modelo deben ser constitutivamente relevantes para el mecanismo que subyace al fenómeno. ¿Cómo debe entenderse esta relación de relevancia constitutiva?

La primera consideración negativa es la siguiente: la relación de constitución no debe entenderse como una relación de super-veniencia. Éstas se dan, en todo caso, entre el mecanismo –considerado como una totalidad– y el fenómeno *explanandum*: no puede haber una diferencia en el fenómeno sin que haya una diferencia en el mecanismo. La relación constitutiva, en cambio, se da entre el mecanismo como totalidad y sus partes componentes, en el contexto de una demanda de explicación.

La segunda consideración negativa es la siguiente: las relaciones constitutivas deben ser netamente distinguidas de las relaciones causales que se dan entre los componentes de un mecanismo, tal como se analizaron en la sección anterior. Se trata de una tesis

metafísica: las relaciones constitutivas entre ítems pertenecientes a distintos niveles de mecanismo no son, ni pueden ser, relaciones causales. El principal argumento de los mecanicistas a favor de esta idea es el que Leuridan (2011, p. 7) denomina “argumento lógico” en contra de la causación internivel¹¹ –aunque se trata más bien de un argumento metafísico.

Argumento en contra de la causación internivel

1. Si es metafísicamente posible la causación entre ítems pertenecientes a distintos niveles de mecanismo (causación internivel), entonces dichos ítems deben ser completamente metafísicamente distintos (principio de Hume-Lewis).
 2. Los ítems pertenecientes a distintos niveles de mecanismo están vinculados por relaciones mereológicas parte-todo.
 3. Si dos ítems están vinculados por una relación parte-todo, entonces no son completamente metafísicamente distintos.
- (C) Por lo tanto, no es metafísicamente posible la causación entre ítems pertenecientes a distintos niveles de mecanismo.

11. En la discusión del “argumento lógico” aparece una distinción entre causación intranivel y causación internivel, respecto de la cual conviene realizar una aclaración preliminar. Básicamente, una relación causal intranivel vincula ítems (objetos, eventos, procesos, etc.) pertenecientes a un mismo “nivel de naturaleza”. Por su parte, una relación causal internivel vincula ítems que pertenecen a distintos niveles de naturaleza. Por ejemplo, la relación causal entre, por un lado, el impacto de una bola A sobre una bola B (en un tiempo *t*) y, por otro lado, el cambio en la aceleración de la bola B (en un instante posterior *t'*) es un caso de causación intranivel. La relación causal entre, por un lado, cierto patrón de activación en un conjunto particular de neuronas del hipocampo y, por el otro lado, la fijación de un recuerdo perceptivo por parte de Juan, es un caso de causación internivel. Puesto que estamos discutiendo el mecanicismo, la noción de “niveles de naturaleza” que resulta relevante aquí es la de *niveles de mecanismo*: la actividad ϕ por parte de *X* pertenece a un nivel inferior respecto de la actividad ψ de *S* si y sólo si la actividad ϕ de *X* es un componente del mecanismo que subyace a la actividad ψ de *S* (Craver 2007, p. 189).

Lo que el argumento explicita es que, si se elucidan las relaciones entre el fenómeno en el nivel de la conducta del sistema como un todo y los componentes del sistema como relaciones mereológicas parte-todo, entonces la causación internivel violaría un principio venerable y ampliamente aceptado acerca de la naturaleza de la causación. La mayoría de los análisis de la causación, al menos desde Hume, han asumido que las causas y los efectos deben ser completamente distintos. El principio en cuestión es formulado por Lewis en los siguientes términos:

C y E deben ser eventos distintos –y no sólo distintos en el sentido de la no identidad, sino en el sentido de la no superposición y la no implicación. No es legítimo decir que el que yo pronuncie esta oración causa el que yo pronuncie esta oración, o que el que yo pronuncie la oración completa causa el que yo pronuncie la mitad de ella; o que el que la pronuncie causa el que la pronuncie en voz alta, o viceversa. (Lewis 2000, p. 78; mis cursivas)

Ahora bien, ¿por qué pensar que las totalidades y sus partes no son completamente metafísicamente distintas? Como Bechtel y Craver (2006) notan (utilizando la terminología de Salmon, aunque el mismo punto podría ser formulado, en principio, en el lenguaje metafísico de su preferencia) las totalidades y sus componentes no son *procesos causales* completamente distintos.

Dadas las relaciones composicionales entre los mecanismos y sus componentes, el curso espaciotemporal de la totalidad incluye el curso espaciotemporal de sus componentes. Ellos coexisten uno con el otro, y no hay posibilidad de que lleguen a intersectarse espaciotemporalmente. (Bechtel y Craver 2006, p. 552).

Un defensor de la causación internivel (por ejemplo, Bermúdez 2000; 2005) podría argumentar que algo tiene que estar mal en el principio de Hume-Lewis, pues existen ejemplos evidentes de explicaciones que requieren postular relaciones causales internivel. En este sentido, Bermúdez (2000) ofrece el llamado “argumento a partir de la conducta habilidosa [*skilled behavior*]”. Supongamos que estoy jugando al tenis y pierdo porque mi oponente realiza *esta* particular volea cuando el partido se encuentra en *match point*. Perder el juego parece ser un caso paradigmático de un evento en el

nivel personal, esto es, asignable a la persona como totalidad. Sin embargo, “para explicar por qué mi oponente extendió su raqueta precisamente esa distancia con ese ángulo preciso” necesitamos contar una historia acerca de cómo el registro, en los sistemas perceptivos de mi rival, de una pelota llegando en un ángulo particular y de mi ubicación en el campo, se coordina con sus sistemas motores para generar configuraciones precisas de movimientos corporales, y éstos últimos “son, por supuesto, hechos subpersonales”, atribuibles a subsistemas o partes de la persona (Bermúdez 2000, p. 80). El ser derrotado en el partido se explica causalmente por las actividades coordinadas de los sistemas (las partes) de la mente/cerebro de mi oponente que están involucrados en la integración perceptiva y el control motor. ¿Qué hace que esta clase de explicación sea causal? Según Bermúdez, satisface el análisis contrafáctico de la explicación causal ofrecido por Schiffer (1991): *F* explica causalmente *G* si, y sólo si: (a) *F* causó *G*; (b) si otro evento *c** hubiese sido el caso en lugar de *F*, y si *c** no hubiese sido un *F*, *c** no habría causado *G*. Si mi oponente no hubiese hecho ese particular arreglo de movimientos corporales, entonces yo no habría perdido el partido de tenis.

Sin embargo, este ejemplo puede ser reinterpretado de manera tal que vuelva ociosa la postulación de relaciones causales internivel. Por supuesto, el ser derrotado en el partido es un evento en el nivel personal. Pero debemos distinguir *el realizar tal conducta habilidosa por parte de mi oponente de las actividades coordinadas de los sistemas involucrados en la integración perceptiva y el control motor*. Lo que causa mi derrota en el juego es que mi oponente haya realizado ese movimiento habilidoso particular. Aquí no hay ningún enigma porque tal realización es también un evento en el nivel personal: la volea es una conducta del jugador como un todo; esos movimientos particulares deben adscribirse a la persona como un todo (como sabiamente lo hacen los asistentes al partido y los comentaristas); es mi oponente como totalidad el que ganó el partido jugando esa volea, y ninguna de sus partes ganó el partido. ¿Por qué mi oponente ganó el partido? Porque jugó esa particular volea. De hecho, existe evidencia neurocognitiva que indica que el mismo arreglo de movimientos habilidosos puede estar realizado en, al menos, dos mecanismos subpersonales distintos: uno involucra el razonamiento consciente, mientras que el otro no (McMorris 2004). Mi derrota

en el juego es explicada *causalmente* por el evento que consiste en el jugar esa volea por parte de mi oponente, pero este último evento no está explicado causalmente, sino constitutiva o componencialmente, por las actividades coordinadas de los subsistemas involucrados en la percepción y el control motor. En el ejemplo, todas las relaciones causales relevantes son intranivel (ya sea en el nivel personal o en el nivel subpersonal) y todas las relaciones internivel no son causales, sino constitutivas. La integración (explícita o implícita) de la información visual y los programas motores no son causalmente relevantes para el fenómeno de la conducta habilidosa: son constitutivamente relevantes para ese fenómeno.

Si esta réplica es correcta, entonces el análisis de las relaciones constitutivas internivel debe distinguir estas últimas de las relaciones causales intranivel. Pues bien, la estrategia de Craver consiste en extender la propuesta manipulacionista de Woodward (2003) y aplicarla al análisis de la relevancia constitutiva. Que esta estrategia sea exitosa no es una cuestión menor: recuérdese que la relevancia constitutiva es la relación metafísica fundamental que vincula el *explanans* y el *explanandum* de una explicación científica. Según Craver:

[MM]

Una variable φ es constitutivamente relevante para otra variable ψ si: (a) φ es una parte de ψ ; (b) ambas variables son mutuamente manipulables. Que sean mutuamente manipulables quiere decir que: (i) existe una intervención internivel ideal I sobre φ , que modifica el valor de la variable ψ y (ii) existe una intervención internivel ideal I^* sobre ψ que modifica el valor de φ .

El concepto fundamental de este análisis es el de *intervención internivel ideal*, que no debe ser confundido con el concepto de intervención etiológica ideal, presentado en la sección anterior. Siguiendo la caracterización de Craver (2007, p. 154), podemos caracterizar el concepto de la siguiente manera:

[C]

Una intervención internivel ideal I sobre φ con respecto a ψ es un cambio en el valor de φ , que cambia ψ sólo vía un cambio en φ . Esto implica lo siguiente: (1) La intervención I no cambia ψ directamente; (2) I no cambia el valor de ninguna otra variable φ^* , que cambia el valor de ψ ,

sino por medio del cambio sobre φ ; (3) I no está vinculada con ninguna otra variable M causalmente independiente de I que sea también una causa de ψ ; (4) I fija el valor de φ de manera tal que filtra o tapa la contribución de las otras causas de φ .

El paralelismo entre las condiciones para una intervención internivel y aquellas de una intervención etiológica es sorprendente y motiva el siguiente problema conceptual (Leuridan 2011, p. 9). Si una intervención ideal satisface los requisitos de (C), entonces satisface también los requisitos de (I). Nótese que si la intervención satisface (C-1)-(C-4), entonces satisface (I-1)-(I-4): basta con reemplazar las variables X e Y , que pertenecen al mismo nivel de mecanismo, por las variables φ e ψ , que pertenecen a niveles de mecanismo distintos. La conclusión es inevitable: si una intervención satisface las condiciones para ser internivel, entonces satisface las condiciones para ser causal.

En este punto de la argumentación, un defensor del mecanicismo enfrenta un dilema: o bien conserva la estrategia manipulacionista respecto de la relevancia constitutiva y ofrece argumentos para bloquear la inferencia al carácter causal de esa relación, o bien abandona el manipulacionismo y elucida la relevancia constitutiva en términos que no sean trivialmente causales.

Quienes optan por el primer cuerno del dilema pueden poner el acento en el hecho de que las relaciones constitutivas no puedan ser causales porque los *relata* son totalidades y sus correspondientes partes. Para que este enfoque funcione debemos contar con una noción clara de *parte* en el contexto del análisis mecanicista. Craver (2006, 2007) ofrece cuatro criterios que nos permitirían distinguir las partes reales de un mecanismo, de las meras “ficciones útiles” o partes ficticias. Son los siguientes: en primer lugar, las partes exhiben un cúmulo estable de propiedades. Esto quiere decir que constituyen clases naturales en el sentido que Boyd (1991) le da al concepto, esto es, un conjunto de propiedades o síntomas que presentan una tendencia a co-ocurrir debido a la acción de un mecanismo homeostático subyacente. En segundo lugar, las partes son robustas, es decir, son detectables mediante una gran variedad de dispositivos experimentales que son causal y teóricamente independientes entre sí. En tercer lugar, un índice de que nos encontramos frente a partes reales de un mecanismo es el hecho de que las

podamos usar para intervenir en otros componentes o actividades del mecanismo (Hacking 1983). Por último, las partes reales deben ser fisiológicamente plausibles. No deberían existir únicamente, por ejemplo, en condiciones de laboratorio totalmente artificiales o en condiciones patológicas.

Considero que, desde el punto de vista de la filosofía de la experimentación en ciencias especiales, el cumplimiento, por parte de un ítem, de los cuatro criterios recién mencionados ofrece fuerte evidencia a favor de la hipótesis de que nos hallamos frente a una *clase natural*, y no meramente frente a un artefacto de nuestros métodos y diseños experimentales. Sin embargo, el que un ítem cumpla estos criterios no nos dice nada acerca de su relevancia como componente de un mecanismo. En otras palabras, el cumplimiento de estos criterios no nos permite distinguir entre los componentes de un mecanismo y las meras condiciones de trasfondo o los efectos estériles. El ejemplo es de Craver (2007). Objetivamente, el latido de mi corazón está conectado, como condición de trasfondo, con mi capacidad para procesar palabras. Si mi corazón dejase de latir, el procesamiento de palabras se vería afectado. No diríamos que mi corazón es un componente constitutivamente relevante de mi mecanismo de procesamiento de palabras. Sin embargo, el corazón cumple con todos los criterios para ser considerado una *clase natural* biológica à la Boyd, y cumple además los otros tres criterios que Craver ofrece para distinguir partes reales de partes ficticias.

La clase de las partes ficticias no es extensionalmente equivalente a la clase de las partes constitutivamente irrelevantes en relación con un mecanismo. Por lo tanto, parece que, aún cuando Craver ofrece criterios para identificar clases reales en general, no ofrece criterios para identificar partes componentes de un mecanismo.

Una última observación: es importante destacar que el análisis de Craver, en términos de la mutua manipulabilidad, se propone ofrecernos sólo condiciones suficientes, pero *no* necesarias, para la relevancia constitutiva. Esto quiere decir que no tenemos criterios para saber cuándo una parte *no* es relevante para un mecanismo. Craver afirma:

Para establecer que la actividad φ de X es *relevante* para la actividad ψ de S , es *suficiente* que uno pueda manipular ψ interviniendo sobre φ (sea por estimulación o por inhibición) y que uno sea capaz de mani-

pular φ manipulando ψ . Para establecer que un componente es *irrelevante*, es *suficiente* mostrar que uno no puede manipular ψ mediante la intervención sobre φ y que uno no puede manipular φ manipulando ψ . (Craver 2007, p. 159)

Sin embargo, esta segunda afirmación está completamente injustificada. La ausencia de relevancia constitutiva es condición suficiente para la ausencia de mutua manipulabilidad, pero la ausencia de esta última no es suficiente para la ausencia de relevancia constitutiva. Este error en el razonamiento de Craver demuestra que, en todo caso, ha ofrecido un “análisis operacional” de la relevancia constitutiva –esto es, ha explicitado qué evidencia es síntoma de esa relación– pero no nos ha dicho en qué consiste, siquiera parcialmente, la naturaleza de esa relación, pues no ha identificado ninguna condición necesaria para ella.

Considérese ahora el segundo cuerno del dilema del mecanicismo. Éste podría intentar ofrecer una elucidación no manipulacionista de las relaciones constitutivas y, por lo tanto, una elucidación no trivialmente causal de estas últimas. Tal es la estrategia que parece adoptar Mark Couch (2011). Según este autor, una parte constitutivamente relevante es *una parte insuficiente, pero no redundante*, de un *mecanismo suficiente, pero no necesario*, que sirve como la realización de una capacidad (Couch 2011, p. 384). Por ejemplo, bajo condiciones normales, en una ocasión particular, los receptores de la retina son necesarios, pero no suficientes, para el procesamiento de la información visual en el ojo humano. Pero el mismo procesamiento podría estar realizado en otro mecanismo distinto del que, de hecho, tenemos; por lo tanto, este último sería suficiente, pero no necesario, para realizar la capacidad.

El lector seguramente habrá notado que esta propuesta parte del análisis que ofrece Mackie (1974) de la causación. Mackie analiza la naturaleza de las relaciones causales en términos de condiciones necesarias y suficientes. Específicamente, un evento causa otro si, y sólo si, el primero es una parte insuficiente, pero no redundante, de una condición suficiente, pero innecesaria, del segundo (condición *inus*). Así, por ejemplo, cuando decimos que un cortocircuito causó un incendio, decimos que el cortocircuito no era una condición suficiente para el incendio (se necesitaba, por ejemplo, que haya oxígeno en el ambiente); sin embargo, era una parte relevante

de la condición que fue suficiente (aunque no necesaria) para que el incendio se produzca. Con miras al análisis de las relaciones constitutivas, Couch afirma:

Supóngase que tenemos una estructura compleja *ABCD* que sirve para realizar una capacidad *F*. Supóngase, además, que, de las partes presentes, sólo *ABC* son necesarias para la presencia de *F* en una ocasión, y que la otra parte *D* es una parte extra que no desempeña ningún papel en la presencia de la capacidad. En este caso, los componentes individuales *A*, *B*, *C* son componentes *inus* para la capacidad *F*, y *D* es una parte irrelevante que debería ser ignorada en la explicación. (Couch 2011, p. 384)

La propuesta de Couch presenta varios problemas conceptuales. En primer lugar, afirmar que una parte relevante es un componente no redundante de un mecanismo no parece ser un análisis muy clarificador; en todo caso, es sólo otro nombre para el problema, pues qué quiere decir que una parte sea no redundante si no quiere decir que, de alguna manera, es una parte relevante. En segundo lugar, está claro que si una parte es constitutivamente relevante en el sentido de Couch, esa parte es causalmente relevante en el sentido de Mackie. Decir, simplemente, que no puede tratarse de una relación causal porque las totalidades no son lógicas ni metafísicamente independientes de sus partes (Couch 2011, p. 385) sólo pone en evidencia la necesidad de un análisis tajantemente no causal de las relaciones de relevancia constitutiva.

5. Conclusión

En este artículo, he defendido la tesis de que los conceptos principales del análisis mecanicista de la explicación científica, aún en las mejores versiones de este último, resultan más oscuros que el concepto mismo de explicación que se pretende elucidar. En otras palabras, el mecanicismo no satisface adecuadamente el requisito de elucidación conceptual que introduje en la sección 2. Recapitularé las razones que apoyan esta afirmación.

En la sección 3 he indicado que los mecanicistas toman el concepto de *modelo científico*, sin mayor justificación, como un primitivo del análisis, y que la única propiedad de los modelos que resulta

relevante para la cuestión de la explicación es la propiedad relacional de *representar* un fenómeno y/o un mecanismo. Este desinterés por las propiedades “intrínsecas” de los modelos científicos les impide ofrecer una caracterización de los modelos fenoménicos que los distinga nítidamente de los modelos ficcionales. Además, los mecanicistas adoptan una concepción de la representación científica en términos de *similitud*, sin ningún escrutinio de los problemas conceptuales asociados a esa concepción. En particular, los mecanicistas no ofrecen una respuesta al “argumento lógico” en contra del análisis de la representación científica en términos de similitud. Mientras la relación de representación es, como mínimo, una relación no reflexiva y no simétrica, la relación de similitud es reflexiva y simétrica. Este problema no se soluciona si se reemplaza la similitud por una relación de isomorfismo, pues esta última es una relación de equivalencia.

En la sección 4 he señalado que, en lo que concierne a la caracterización del fenómeno *explanandum* de un modelo mecanicista, estos autores adoptan un principio (la “ley de Glennan”) según el cual no puede haber, ni hubieron, explicaciones mecanicistas de fenómenos inexistentes o ficticios, ni de fenómenos que puedan dividirse en otros fenómenos, a cada cual correspondiéndole un mecanismo distinto. Este principio es demasiado restrictivo, pues resulta plausible pensar que pueden ofrecerse (y que se han ofrecido) modelos “mecanicistas” (en un sentido relevante) de fenómenos que luego resultaron ser inexistentes o que no estaban del todo comprendidos en su momento. Conscientes de este problema, muchos mecanicistas reemplazan el principio en cuestión por la afirmación (casi trivial) de que los modelos mecanicistas y la comprensión de los fenómenos “evolucionan conjuntamente”. Esta posición resulta aceptable justamente porque no establece una restricción sobre qué tipos de fenómenos son pasibles de una explicación mecanicista.

De manera crucial, en “Las partes constitutivamente relevantes” de la sección 4, he argumentado que los mecanicistas no nos han ofrecido aún un análisis satisfactorio de la relación explicativa fundamental del análisis, esto es, de la relación de relevancia constitutiva que vincula el mecanismo como totalidad y sus partes componentes. Por un lado, si se analiza la relación de relevancia constitutiva en términos de mutua manipulabilidad, entonces aquella resulta (implícitamente) causal. Pero existen buenas razones meta-

físicas –aceptadas por los propios mecanicistas– para pensar que no puede haber relaciones causales entre ítems pertenecientes a distintos niveles de mecanismo, como es el caso de los *relata* en una relación constitutiva. Por otro lado, si se analiza la relación de relevancia constitutiva en términos de condiciones necesarias y suficientes (en analogía con el análisis *inus* de las relaciones causales), entonces la relación de relevancia constitutiva resulta (casi explícitamente) causal. En cualquier caso, los análisis disponibles de las relaciones de relevancia constitutiva no logran distinguir éstas últimas de las relaciones de relevancia causal.

En suma, considero que el entusiasmo por la concepción mecanicista de la explicación debe moderarse hasta tanto sus defensores enmienden los problemas conceptuales que se encuentran en el núcleo de su propuesta. Mientras tanto, no parece haber buenas razones para pensar que el mecanicismo satisfaga, de manera mínimamente aceptable, el requisito de elucidación conceptual.¹²

BIBLIOGRAFÍA

- BARROS, B. (2008) "Natural Selection as a Mechanism", *Philosophy of Science* 75: 306-322.
- BARTELBORTH, T. (1996) "Scientific Explanation", en W. Balzer & C.U. Moulines (eds.) *Structuralist Theory of Science: New Results*, Berlin, de Gruyter, 23-44.

12. Este trabajo fue concebido y elaborado en el marco del Grupo de Investigación en Cognición, Lenguaje y Conciencia (CLC), coordinado por Liza Skidelsky y Diana Pérez. Le debo mi más profundo agradecimiento a Liza Skidelsky por su estímulo constante, por su calidez humana y por la agudeza de sus observaciones a cada una de mis producciones filosóficas. Agradezco a los miembros del Grupo CLC, especialmente a Sabrina Haimovici y Mariela Destéfano. Una versión previa de este trabajo fue presentada en el Grupo de Acción Filosófica. Agradezco a los participantes de dicha reunión, especialmente a Eleonora Cresto, no sólo por la generosidad de su invitación sino también por sus comentarios sobre mi presentación. Por último, quisiera agradecer a Alejandro Cassini, quien amablemente leyó un esbozo de este trabajo y me animó a que lo convierta en la presente publicación.

- BECHTEL, W. (2008) *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*, New York, Routledge.
- BECHTEL, W. & ABRAHAMSEN, A. (2005) "Explanation: a Mechanist Alternative", *Studies in the History and Philosophy of Biological and Biomedical Sciences* 36: 421-441.
- BECHTEL, W. y RICHARDSON, R. (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton, Princeton University Press.
- BECHTEL, W. & WRIGHT, C. (2007) "What is Psychological Explanation" en P. Calvo & J. Symons (eds.) *Routledge Companion to the Philosophy of Psychology*, New York, Routledge, 113-130.
- BERMUDEZ, J. (2000) "Personal and Subpersonal: A Difference without a Distinction", *Philosophical Explorations*, 3(1): 63-82.
- BERMÚDEZ, J. L. (2005) *Philosophy of Psychology: A Contemporary Introduction*, New York, Routledge.
- BOGEN, J. (2005) "Regularities and Causality; Generalizations and Causal Explanations", *Studies in History of Philosophy of Biology and Biomedical Sciences* 36: 397-420.
- BOYD, R. (1991) "Realism, Anti-foundationalism and the Enthusiasm for Natural Kinds", *Philosophical Studies* 61: 127-148.
- CHURCHLAND, P. (1989) *A Neurocomputational Perspective*, Cambridge, Bradford/MIT Press.
- COUCH, M. (2011) "Mechanisms and Constitutive Relevance", *Synthese* 183: 375-388.
- CRAVER, C. & BECHTEL, W. (2007) "Top-Down Causation without Top-Down Causes", *Biology and Philosophy* 22: 547-563.
- CRAVER, C. (2002) "Structures of Scientific Theories", en P. Machamer & M. Silberstein (eds.) *Blackwell Guide to the Philosophy of Science*, Oxford, Blackwell, 55-79.
- CRAVER, C. (2006) "When Mechanistic Models Explain", *Synthese*, 155, pp. 355-376.
- CRAVER, C. (2007) *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford, Clarendon Press.
- CUMMINS, R. (1983) *The Nature of Psychological Explanation*, Cambridge, MIT Press.
- DARDEN, L. (2006) *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations and Anomaly Resolution*, Cambridge, Cambridge University Press.
- DÍEZ, J. (2002) "Explicación, unificación y subsunción", en W. Gon-

- zález (ed.) *Pluralidad de la Explicación Científica*, Ariel, Barcelona, 73-93.
- DÍEZ, J. (2008) "La explicación científica: causalidad, unificación y subsunción teórica", en L. Hoyos (ed.) *Relativismo y Racionalidad*, Bogotá, UNC, 283-414.
- FODOR, J. (1974) "Special Sciences and the Disunity of Science as a Working Hypothesis", *Synthese* 28: 77-115.
- FRIGG, R. (2006) "Scientific Representation and the Semantic View of Theories", *Theoria* 55: 37-53.
- GIERE, R. (1999) *Science without Laws*, Chicago, University of Chicago Press.
- GLENNAN, S. (2002) "Rethinking Mechanistic Explanation", *Philosophy of Science* 69: 342-353.
- GLENNAN, S. (2005) "Modelling Mechanisms", *Studies in the History and Philosophy of Biological and Biomedical Sciences* 36: 443-464.
- GODFREY-SMITH, P. (2006) "The Strategy of Model-Based Science", *Biology and Philosophy* 21: 725-740
- GOODMAN, N. (1976) *Languages of Art: An Approach to a Theory of Symbols*, Indianapolis, Hackett Publishing Company.
- HACKING, I. (1983) *Representing and Intervening*, Cambridge, Cambridge University Press.
- HEMPEL, C. (1965) *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, New York, The Free Press.
- HODGKIN, A. y HUXLEY, A. (1952) "Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve", *Journal of Physiology* 117: 500-544.
- KAPLAN, D. M. (2011) "Explanation and Description in Computational Neuroscience", *Synthese* 183: 339-373.
- KITCHER, P. (1981) "Explanatory Unification", *Philosophy of Science* 48(4): 507-531.
- LEURIDAN, B. (2011) "Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms", *British Journal of Philosophy of Science*, online.
- LEWIS, D. (2000) "Causation as Influence", *Journal of Philosophy* 97: 182-197.
- MACHAMER, P., DARDEN, L. & CRAVER, C. (2000) "Thinking about Mechanisms", *Philosophy of Science* 67: 1-25.
- MACKIE, J. (1974) *The Cement of the Universe: A Study of Causation*, Oxford, Clarendon Press.

- MATTHEWSON, J. & CALCOTT, B. (2011) "Mechanistic Models of Population-Level Phenomena", *Biology and Philosophy* 27: 737-756.
- MCMORRIS, T. (2004) *Acquisition and Performance of Sports Skills*, Chichester, Wiley.
- MORGAN, M. y MORRISON, M. (1999) *Models as Mediators*, Cambridge, Cambridge University Press.
- NAGEL, E. (1961) *The Structure of Science: Problems in the Logic of Scientific Explanation*, Harcourt, Brace & World.
- PICCININI, G. & CRAVER, C. (2011) "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches", *Synthese* 183: 283-311.
- RUMELHART, D. y MCCLELLAND, J. (1986) *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, Cambridge, MIT Press.
- SALMON, W. (1984) *Scientific Explanation and the Causal Structure of the World*, Princeton, Princeton University Press.
- SALMON, W. (1998) *Causality and Explanation*, Oxford, Oxford University Press.
- SCHIFFER, S. (1991) "Ceteris Paribus Laws", *Mind* 100: 1-17
- SUÁREZ, M. (2003) "Scientific Representation: Against Similarity and Isomorphism", *International Studies in the Philosophy of Science*, 17(3): 225-244.
- VAN FRAASSEN, B. (1980) *The Scientific Image*, Oxford, Clarendon Press.
- WEISKOPF, D. (2011) "Models and Mechanisms in Psychological Explanation", *Synthese* 183: 313-338.
- WOODWARD, J. (2002) "What is a Mechanism? A Counterfactual Account", *Philosophy of Science* 69: 366-377.
- WOODWARD, J. (2003) *Making Things Happen: a Theory of Causal Explanation*, Oxford, Oxford University Press.
- WRIGHT, C. (inédito) "Scientific Explanation: Mechanistic, Ontic, Epistemic".

Recibido: 05-2011; aceptado: 08-2012