Quaderns de filosofia i ciència, 37, 2007, pp. 7-14

EXPLICACIÓN TEÓRICA Y COMPROMISOS ONTOLÓGICOS: UN MODELO ESTRUCTURALISTA

C. Ulisses Moulines
Seminar für Philosophie, Logik und Wissenschaftstheorie
Universitat de Munic

Abstract: We try to present some formal models of explanation, with no universal validity but with a wide range of applications. It will be done with a particular kind of *explicandum*, to be called "theoretical explanation with ontological commitments", which has a putative significance in many sections of the natural sciences, mainly in physics, and in those disciplines that have been mathematized in a systematic way.

Keywords: scientific explanation, formal model, ontological commitment, structuralism, scientific theory.

Más de cincuenta años de discusiones bastante frustrantes alrededor del concepto de *explicación científica* (para no hablar del concepto de explicación *tout court*) indican en mi opinión claramente que no podemos esperar que se desentrañe una estructura subyacente única para todo tipo de explicaciones en los diferentes contextos científicos. Sin embargo, sería un *non sequitur* inferir de esta situación que debemos abandonar cualquier esperanza de alcanzar cierta comprensión de los modos y usos de la explicación científica, o que debamos refugiarnos en un "todo vale" feyerabendiano. Lo que, en mi opinión, podemos y debemos hacer es ser más prudentes y atrevidos a la vez, y ello significa aquí idear ciertos modelos formales, controlables, de lo que puede ser una explicación, modelos para los que no pretenderemos de entrada que tengan una validez universal, pero que pueden tener un espectro de aplicaciones más o menos amplio. Esto es lo que me propongo hacer en este ensayo para un tipo particular de *explicandum* que, a falta de mejor rótulo, podemos denominar "explicación teórica con compromisos ontológicos". Concedo de inmediato que este tipo de explicación apenas tiene algo que ver con el uso del término *explicación* en la vida cotidiana, y que no es ni siquiera aplicable a todos los contextos científicos. Pero pretendo que tiene una significación genuina en grandes porciones de las ciencias naturales, particularmente en la física, y más generalmente en aquellas disciplinas que han sido matematizadas de una manera más o menos sistemática. La aplicabilidad del modelo de explicación que se propone aquí presupone, no obstante, que aceptemos algunas premisas generales, si se quiere *metafísicas*, sobre la ciencia. Estas premisas son:

1. La ciencia consiste, al menos en parte, de cosas llamadas *teorías*.
2. Las teorías pueden ser identificadas por medios modelo-teóricos.
3. Las teorías se usan, entre otras cosas, para explicar los fenómenos.

4. Al menos, algunas porciones de la ciencia teórica (es decir, de aquella parte de la ciencia que consiste en teorías) permiten una interpretación realista; ello significa que los compromisos ontológicos hechos en el seno de estas teorías deben ser tomados literalmente en cuanto que refiriéndose a "las cosas ahí fuera" (o sea, cosas independientes de la comunidad de sujetos epistémicos que usa las teorías).
5. *Puede* que haya *causas* de los fenómenos observados (o, más precisamente, *puede* que alguna noción adecuada de *causa* sea relevante para dar cuenta de la explicación teórica de los fenómenos, al menos en ciertos casos).

Soy perfectamente consciente de que ninguno de estos supuestos es incontrovertido. Un filósofo "experimentalista" al estilo de Hacking o Cartwright (al igual que un operacionalista radical, por cierto) tendrá problemas con el supuesto 1, y en consecuencia también con la tesis 2; instrumentalistas "abiertos" a lo Duhem o "encubiertos" a lo van Fraassen rechazarán 3; los anti-realistas de toda índole (incluyendo a los instrumentalistas) negarán enfáticamente 4; y finalmente muchas clases de positivistas, empiristas e instrumentalistas se sonreirán ante la suposición 5.

No voy a embarcarme en una argumentación a favor de estos supuestos. Por otro lado, me parece obvio que la mayoría de los científicos practicantes, al menos en la medida en que aún no se han visto involucrados en discusiones filosóficas, tomarán los "axiomas metateóricos" 1 a 5 como algo incontrovertible, si se les pregunta. ¿Superstición de los científicos? Quizá. Pero, dado que creo que una tarea de primer orden para el filósofo de la ciencia consiste en explicitar los supuestos implícitos de los científicos, voy a tomar dichas premisas como mi punto de partida y a preguntarme qué resulta de ellas para el tema de la explicación científica.

Mi tesis general, que voy a tratar de articular en lo que sigue, es simplemente la siguiente: al menos en muchos contextos de la ciencia teórica, la explicación adopta la forma: se *inserta* una estructura de datos en un modelo teórico, algunos de cuyos componentes, precisamente los teóricos, puede que admitan una interpretación causal. Esta tesis no es altamente original. Puede ser considerada como una especie de síntesis del enfoque unificacionista de Friedman y Kitcher con el enfoque causalista de Wesley Salmon, modificada y enriquecida por desarrollos más recientes debidos sobre todo a algunos autores que trabajan dentro desde la perspectiva del estructuralismo metateórico, tales como Thomas Bartelborth, John Forge y José A. Díez. De hecho, estos enfoques tampoco son incompatibles con la consideración de factores pragmáticos en las explicaciones científicas, a la manera como lo propone van Fraassen. Todo lo que pretendo hacer en lo que sigue es ensamblar estas piezas y añadirles un poco de sal y pimienta de mi parte; esto último consistirá principalmente en el énfasis puesto en las nociones de *modelos de datos* y de *compromisos ontológicos* "reales".

En lugar de embarcarme en una argumentación general, voy a proceder de manera wittgensteineana y empezar con una especie de *Gedankenexperiment* bajo la forma de un ejemplo simple, esquemático.

Supongamos que Pepe (o su ancestro babilónico Nabucodonosor) no tiene nada más que hacer sino yacer en su terraza durante una noche de insomnio, contemplando el cielo estrellado. De esa manera, se ve enfrentado a una "situación experiencial", llamémosla *SE*, que le intriga y que podemos describir como sigue: "experiencia del cielo nocturno en una noche clara; puntos luminosos en diferentes posiciones que se mueven lentamente". Pepe-Nabucodonosor quiere entender qué es lo que está pasando, o sea *explicárselo*, como quizá también diríamos en este contexto. ¿Cuál podemos imaginar que será su siguiente paso?

Si Pepe-Nabucodonosor adopta una actitud realmente científica, no se mostrará egocéntrico, sino que se reunirá con un grupo de personas que también se sientan intrigadas por las extrañas cosas que están pasando en el cielo nocturno, y todos juntos constituirán un "grupo de socios" (llamémosle *GS*), dispuestos a cooperar para averiguar qué está pasando. Ellos "negociarán", como dirían los sociólogos de la ciencia, un procedimiento de investigación que consistirá básicamente en los siguientes pasos:

Primero, se ponen de acuerdo para llevar a cabo sus observaciones durante un gran número de noches sucesivas.

Segundo, se ponen de acuerdo en determinar los intervalos de tiempo entre las sucesivas posiciones de cada uno de los puntos luminosos mediante un instrumento llamado *reloj*. (No tenemos por qué imaginar que se trata aquí de un aparato de alta tecnología; basta un reloj de arena, por ejemplo, sobre el cual se han hecho ciertas marcas regulares.)

Tercero, *GS* decide concentrarse en aquellos puntos luminosos que muestran un movimiento irregular durante una larga sucesión de noches, empezando siempre a partir de la misma marca sobre el reloj de arena y tomando los mismos intervalos (o sea, las mismas distancias entre las marcas sobre el reloj).

Cuarto, *GS* decide darles a los puntos luminosos que se mueven irregularmente un nombre genérico, por ejemplo, *planetas*, y cada uno de ellos recibe también un nombre individual: *Mercurio*, *Venus*, etc.

Hasta aquí, *GS* no tiene ninguna teoría, ni ninguna explicación, ni ningún compromiso ontológico en absoluto. Nada ha sido explicado y nada se ha dicho acerca de qué es lo que los puntos luminosos "realmente son". Lo que constatamos es la transformación de la situación experiencial original *SE* que tenía Pepe/Nabucodonosor en una "situación experiencial controlada intersubjetivamente", llamémosla *SECI*.

Ahora puede empezar una nueva fase en la empresa científica dedicada a los puntos luminosos. *GS* se da cuenta de que, para proseguir sus investigaciones sobre el cielo nocturno de manera concienzuda, las meras observaciones y denominaciones no son suficientes. Hay que "fijarlas sobre el papel". Si se me permite una formulación un poco altisonante, podemos decir que lo que necesita *GS* es una *representación* apropiada, controlable, de la experiencia codificada por *SECI*. *GS* decide *representar SECI* de la siguiente manera: toman papel cuadriculado, se ponen de acuerdo en fijar un *centro de coordenadas* que representa un punto luminoso particular (por ejemplo, el punto denominado *Estrella Polar*) y representan la posición relativa de los planetas con respecto a la Estrella Polar dentro de intervalos regulares mediante marcas sobre el papel cuadriculado. Sobre un gran número de hojas de papel se obtiene así un gran número de puntos. Llamaremos a esta representación un *modelo de datos correspondiente a SECI*, o simplemente *MD*.

El modelo de datos de este ejemplo puede sintetizarse por medio de la siguiente estructura:

$$MD = < \{p_1, ..., p_5\}, \{\ldots m_i \ldots\}, d >,$$

donde los p_i representan los puntos luminosos que se están investigando (por ejemplo p_1 es la abreviación del nombre propio *Mercurio*, etc.), m_i representan las marcas sobre el reloj de arena (que podemos representar, por ejemplo, mediante números enteros), y d es una función que consiste de los triplos

$$< p_i, m_i, r_i >,$$

donde los r_i pertenecen a los números racionales y representan las distancias determinadas sobre el papel cuadriculado.

Pero seguimos sin tener ninguna teoría, ninguna explicación, ninguna ontología. Nada se ha dicho acerca de qué son realmente los puntos luminosos, ni acerca de cómo es que toman las posiciones que toman.

Supongamos ahora que alguien (ya sea el propio *GS* o bien alguien distinto) ha construido una teoría particular, llamémosla *T*, para ayudar a *GS* en su empresa, o simplemente para divertirse. De momento, *T* sólo es una entidad matemática. Para identificarla formalmente, asumamos aquí la siguiente premisa metodológica común a todas las concepciones semánticas de la actual filosofía de la ciencia, y característica en particular del estructuralismo metateórico: la mejor manera de elucidar la estructura básica de cualquier teoría científica *T* consiste en identificar la clase de sus modelos *M[T]*. En términos generales, los elementos de una tal clase tendrán la siguiente forma:

$$x = < D_1,..., D_k, A_1,..., A_l, R_1, ..., R_m, f_1, ..., f_n >,$$

donde los D_i (para $i > 0$) son conjuntos simples de entidades no ulteriormente analizadas, los A_i (para $i \geq 0$) son conjuntos "auxiliares" (típicamente espacios numéricos o vectoriales), las R_i (para $i \geq 0$) son relaciones definidas sobre algunos D_i y eventualmente algunos A_i y las f_i (para $i \geq 0$) son funciones definidas sobre algunos D_i y eventualmente algunos A_i.

En nuestro ejemplo, los elementos *x* de *M[T]* puede que tengan la forma:

$$< P, Z, \text{IR}, f_1, ..., f_n >,$$

donde los componentes de cada *x* vendrán caracterizados como sigue:

(T1) *P* es un conjunto finito, no-vacío (cuyos elementos son denominados simplemente *puntitos* por el inventor de *T*).
(T2) *Z* es isomorfo a un intervalo de IR (a cuyos elementos se les llama *instantes*, como se les podría llamar cualquier otra cosa).
(T3) Las f_i son funciones definidas sobre *P* y/o *Z* y/o IR.

Supongamos además que cada $x \in M[T]$ satisface ciertos "axiomas propios" o "leyes", es decir, fórmulas que vinculan entre sí *P, Z* y f_i. (A modo de ilustración, estas fórmulas podrían tener la forma de las leyes de Kepler, o las de Newton, o algo parecido.)

Ahora disponemos de una teoría, pero aún no tenemos ni explicaciones ni compromisos ontológicos "serios". *T* es simplemente un "juego con símbolos", y sus compromisos ontológicos se limitan, en el mejor de los casos, a los números reales, no se refieren a nada en la "realidad externa". Ser un puntito, por el momento, significa solamente ser un elemento de *P*, el cual, a su vez, no es más que el primer componente de una estructura particular *x* que es elemento de *M[T]*; lo mismo vale, naturalmente, para "instante".

Supongamos, no obstante, que mientras *GS* sigue preocupándose por los puntos luminosos llamados *planetas*, a alguien se le ocurre la siguiente idea crucial: por las razones que sean, *GS* empieza a pensar que *T* podría tener algo que ver con *MD*. En la jerga específica del

estructuralismo metateórico diríamos que *GS* concibe ahora *MD* (y de manera derivativa la *SECI* original) como una *aplicación intencional* de *T*.

Supongamos finalmente que algún miembro de *GS* logra demostrar matemáticamente la siguiente aserción, que puede considerarse como una versión generalizada, modelo-teórica de lo que en la filosofía clásica de la ciencia se denomina *un enunciado de Ramsey*:

[R] Para algún $a^\circ \in$ M[T]: MD *es una subestructura de* a°.

El significado preciso de *ser una subestructura* lo elucidamos como sigue:

$$\text{Sea } a^\circ = < P^\circ, Z^\circ, \text{IR}, f^\circ_1, ..., f^\circ_n >$$

donde podemos escoger una de estas f°_i, rebautizarla con el símbolo s° y caracterizarla como sigue:

$$s^\circ : P^\circ \times Z^\circ \mapsto \text{IR}^2.$$

Entonces la fórmula "$MD \prec a^\circ$" (a ser leída: "*MD* es una subestructura de a°") significa que se cumplen las siguientes condiciones:

(1) $\{p_1, ..., p_s\} \subseteq D^\circ$;

(2) $\{..., m_i, ...\} \subseteq Z^\circ$;

(3) $s^\circ / \{p_1, ..., p_s\} \times \{..., m_i, ...\} \times \text{IR}^2 = d$.

Sobre la base del enunciado *[R]*, del cual suponemos que ha sido demostrado matemáticamente, es legítimo que *GS* haga ahora una inferencia ulterior, realmente sustancial:

[S] El modelo de datos MD*, y en un sentido derivativo la experiencia original* SECI*, pueden subsumirse bajo la teoría* T.

Está claro que esto no es un "teorema" en el sentido usual, es decir, una proposición que pueda deducirse formalmente de *[R]* con los medios formales que *GS* tiene a su disposición. Se trata más bien de una inferencia cuasi-analítica basada en la idea que *GS* tiene de lo que significa que una situación experiencial particular pueda subsumirse bajo una teoría. En términos estructuralistas generalizados, *[S]* corresponde a la aserción de que *MD*, y en un sentido derivativo la propia *SECI*, es una "aplicación intencional *exitosa*" de *T*. Es sólo a partir de este punto que *GS* puede estar dispuesto a extraer capital explicativo y ontológico de la situación descrita. Si aceptamos todos los pasos que sucesivamente han conducido de la construcción inicial de *SECI* hasta la justificación de la aserción *[S]*, creo que también estaremos justificados en establecer las dos aserciones siguientes:

[O] Los puntos luminosos llamados planetas *son en realidad* puntitos *(en el sentido de la teoría* T*).*

Y además

[E] T *explica por qué los planetas se comportan como lo hacen.*

Pues bien, mi tesis es que el modelo de explicación ilustrado mediante este ejemplo esquemático es aplicable a, por lo menos, una gran porción de la ciencia teórica, en particular a aquellas partes de la ciencia que han sido más o menos bien matematizadas, aunque no lo es necesariamente a cualquier contexto científico en el que la gente hable de explicación. Podemos denominar a este esquema *la versión débil de la explicación por medio de la subsunción modelo-teórica*. Esta versión todavía es débil en el siguiente sentido: hasta aquí, nada en este modelo de explicación contiene algún elemento que pueda interpretarse como rasgos *unificatorios*, ni tampoco *causales* de la explicación. Esto es, el modelo subsuntivo de la explicación expuesto hasta aquí es compatible con aquellos enfoques de la explicación científica que consideran que ésta no tiene nada que ver ni con la idea de unificación ni tampoco con relaciones causales. El modelo propuesto hasta aquí es muy general, y en este respecto, débil.

Ahora bien, la situación cambia sustancialmente si estamos dispuestos a aceptar toda la batería de nociones estructuralistas para una representación adecuada de la estructura esencial de la ciencia teórica. Algunos componentes de la representación estructuralista de las teorías científicas tienen una interpretación directamente *unificacionista*; éstos son: primero, la idea de que las teorías científicas normalmente tienen la estructura de una *red ordenada jerárquicamente* de elementos teóricos; segundo, la idea de la presencia de *condiciones de ligadura (constraints)* en tanto que principios-puente entre los diversos modelos de una misma teoría, y, finalmente, la idea de la presencia de *vínculos interteóricos (links)* en tanto que principios-puente entre los modelos de teorías diferentes. Además, aunque de una manera menos directa, la idea de que el marco conceptual en la mayoría de las teorías permite una distinción entre dos niveles metodológicos –el nivel de los *conceptos T-teóricos* y el nivel de los conceptos *T-no-teóricos*–, puede interpretarse en el sentido de la introducción de una estructura *causal* en la aserción empírica de una teoría. Empecemos por los elementos que conducen a una interpretación unificacionista de la explicación.

En primer lugar, cuando un modelo teórico $a°$ del elemento teórico $T°$ subsume cierto modelo de datos $MD°$, hay que tomar en cuenta que $T°$ es parte de toda una red teórica $N°$ ordenada por la relación de *especialización*. Ello significa que el proceso de subsunción debe tomar en cuenta el hecho de que las leyes específicas de $T°$ satisfechas por $a°$ implícitamente presuponen la validez de leyes más generales características de los elementos T_i de la red supra-ordenados a $T°$ y por lo tanto a $a°$. En consecuencia, en cualquier subsunción de una porción particular de la experiencia bajo un modelo teórico dado está involucrado explícita o implícitamente todo un pedazo de una red teórica. Si la aserción-Ramsey *[R]* fuera verdadera para $T°$ pero no para los elementos teóricos supra-ordenados a $T°$, entonces no aceptaríamos la explicación de $MD°$ por medio de $a°$. Esta es la reconstrucción estructuralista formal de una parte del enfoque holista o unificacionista de Duhem (y de Friedman y Kitcher).

En segundo lugar, incluso dentro del propio elemento teórico considerado $T°$, $a°$ estará habitualmente conectado por medio de las llamadas *ligaduras* a modelos diferentes que subsumen otras porciones de la experiencia. De manera análoga a las consideraciones previas, la explicación de $MD°$ por medio de $a°$ al modo de *[R]* no se aceptará si se ignoran

esas ligaduras. Puede que las ligaduras involucren incluso la totalidad de la red teórica *N°*. Obviamente, ello tiene consecuencias unificacionistas.

En tercer lugar, *a°* usualmente estará conectado no sólo con otros modelos de la misma red teórica *N°* por medio de las ligaduras, sino también con otros modelos de diferentes redes teóricas *N', N''*,…, por medio de los llamados *vínculos interteóricos*. Una vez más, esto es claramente un rasgo unificatorio del proceso de subsunción de una porción de la experiencia bajo el modelo de una teoría.

Todos estos elementos explicativos adicionales que deben tomarse en consideración según el estructuralismo, pueden formalizarse perfectamente en términos modelo-teóricos, como sabe cualquiera que esté familiarizado con lo esencial de la representación estructuralista de la ciencia. El enunciado de Ramsey modificado que resulta de todo ello, llamémoslo *[RS]*, aparece como un enunciado mucho más complejo que el original *[R]*… pero qué remedio, así es la vida (científica). En cualquier caso, *[RS]* puede formularse en términos bien precisos. Por razones de espacio (y porque el lector puede encontrar la forma general de *[RS]* en la literatura pertinente[1]), no proporcionaré esta formalización aquí.

Para concluir, paso a aquel aspecto de la representación estructuralista de las teorías que puede ser interpretado *causalmente* en el sentido de Wesley Salmon (con cierta buena voluntad). *Si* aceptamos que tiene sentido hacer una distinción entre conceptos *T*-teóricos y conceptos *T*-no-teóricos dentro de una teoría *T*, y *si* admitimos una interpretación *realista* de los conceptos *T*-teóricos, *entonces* es plausible considerar las entidades a las que se refieren los conceptos *T*-teóricos como *causas* del comportamiento específico de *SECI*. Las entidades teóricas a las que se refieren los conceptos *T*-teóricos se introducirían así para proporcionar una explicación causal de los procesos fenoménicos codificados en nuestros modelos de datos. Ellas serían parte de la estructura causal "oculta" del mundo que es responsable de los fenómenos que observamos y codificamos por medios *T*-no-teóricos. En su artículo "Reflections on Structuralism and Scientific Explanation" (*Synthese*, 2002, 130), John Forge sugirió una idea análoga, al menos para el caso de la mecánica newtoniana: según Forge, en este ejemplo, el concepto *T*-teórico de *fuerza* se introduce para identificar una causa de la cinemática particular de un sistema de partículas, estando esta última descrita en términos puramente *T*-no-teóricos. No obstante, Forge presenta su idea de manera muy cautelosa y hace observar que la función de los conceptos *T*-teóricos no siempre consiste en señalar las causas de una situación descrita *T*-no-teóricamente. El supuesto contraejemplo que él presenta es el de la termodinámica: allí, la entropía es claramente un concepto *T*-teórico, pero nadie pretendería afirmar que los cambios en la entropía de un sistema son causalmente responsables de los cambios que sufre el sistema en su volumen o en su temperatura, pongamos por caso. Sin embargo, por mi parte, objetaría a este supuesto contraejemplo que si *hoy día* nos resistimos a ver en la entropía la causa del comportamiento fenoménico de los gases, ello es porque estamos convencidos (por buenas o malas razones) de que la totalidad de la termodinámica, incluyendo sus conceptos *T*-teóricos como el de *entropía*, es *reducible* a una teoría diferente, a saber, la mecánica estadística, y que, por consiguiente, las causas genuinas del comportamiento de los gases deben buscarse en un marco teórico mecánico-estadístico. Pero esto es un problema distinto. La introducción de relaciones interteóricas tales como la reducción en la discusión de la explicación causal complicaría aún más el análisis y no puedo tratar esta cuestión aquí. Me limitaré a observar

[1] Por ejemplo, en W. Balzer, C. U. Moulines y J. D. Sneed, *An Architectonic for Science*, 1987.

que, si no consideramos esta complicación ulterior, no veo ninguna razón (si somos realistas respecto a los conceptos *T*-teóricos) por la cual deberíamos ser tan cautelosos, y que ese exceso de cautela nos conduzca a restringir la interpretación causal de dichos conceptos a solamente algunos de ellos. En la medida en la que admitamos que los conceptos *T*-teóricos usualmente se refieren a algo externo a las teorías en las que aparecen, y que las relaciones causales son parte de la realidad, entonces yo abogaría por una interpretación causal de *todos* los conceptos *T*-teóricos.

Así pues, podemos sintetizar los componentes de lo que puede llamarse *la versión fuerte de la explicación por medio de la subsunción modelo-teórica* en los siguientes términos: explicar una situación experiencial controlada intersubjetivamente consiste, primero, en codificarla en un modelo de datos, y luego, en subsumirla bajo un modelo teórico de una red teórica que satisfaga cierto número de ligaduras intermodélicas y vínculos interteóricos, y haciendo esto último de tal manera que los conceptos teóricos específicos de la teoría en cuestión se interpreten causalmente.