

Manipulando sombras. Hacia una (di)solución evidencialista del problema de la exclusión causal

LORENZO BARAVALLE

§1. Introducción

UNO DE LOS MÁS PERSISTENTES DESAFÍOS en filosofía de la mente y en las filosofías de las ciencias especiales es, sin duda, el conocido problema (o argumento) de la exclusión causal (Kim 1998, 2005). El problema puede ser resumido de la siguiente manera

Si aceptamos:

- el cierre causal de la física;
- la no-identidad de las propiedades supervenientes con sus realizaciones físicas;

y rechazamos, como sumamente implausible, la posibilidad de que los fenómenos físicos sean sistemáticamente sobredeterminados;

entonces tenemos que admitir que las realizaciones físicas son causalmente suficientes para sus efectos y que las propiedades supervenientes son causalmente inertes.

El blanco de la exclusión causal, en las intenciones de Kim, era la causación mental; sin embargo (y no obstante Kim niegue eso; cf. Kim 1998), puede argumentarse que el problema se generaliza a cualquier propiedad múltiplemente realizable (Marras 1999; Bontly 2002). Eso implicaría, en última instancia, que cualquier disciplina diferente de la física sería incapaz —en la medida en que describe relaciones entre propiedades múltiplemente realizables— de formular atribuciones causales verdaderas.

Ciertamente, esta es una consecuencia indeseable. Biólogos, psicólogos y

L. Baravalle (✉)
Universidade de Lisboa, Portugal
e-mail: lbaravalle@fc.ul.pt

Disputatio. Philosophical Research Bulletin
Vol. 8, No. 11, Dec. 2019, pp. 355-380
ISSN: 2254-0601 | [SP] | **ARTÍCULO**

sociólogos hablan constantemente de relaciones causales entre variables y sería, de verdad, muy sorprendente que estuvieran expresándose erróneamente todo el tiempo. Recientemente, un cierto número de autores —ocasionalmente apodados de «evidencialistas» (Baumgartner 2017)— han argumentado que la teoría manipulacionista, o intervencionista, de la causación (Woodward 2003) tiene recursos conceptuales suficientes para bloquear el argumento de la exclusión causal (Shapiro & Sober 2007; Raatikainen 2010; Weslake *forthcoming*; Woodward 2015, 2018).

De acuerdo con el manipulacionismo, o intervencionismo, dos propiedades —o, mas en general, dos variables— X y Y están causalmente relacionadas si y solamente si existen posibles manipulaciones de X tales que, como resultado de tales manipulaciones, X genera cambios correspondientes en Y. Dado que, aparentemente, muchas propiedades supervenientes satisfacen esta condición, los evidencialistas concluyen que no hay razón para denegarles eficacia causal.

Michael Baumgartner (2009, 2010, 2013, 2017) ha levantado una serie de objeciones en contra de esta solución al problema de la exclusión causal. Dos de ellas son especialmente relevantes. La primera objeción (Baumgartner 2009, 2010) concierne la caracterización original de Woodward (2003) de lo que vale como manipulación *posible*. Como discutiré en detalle más adelante, Baumgartner niega que, en el sentido en que Woodward define la noción de manipulación, sea posible manipular propiedades supervenientes. La segunda objeción (Baumgartner 2013, 2017) es que, aunque fuera posible manipular propiedades supervenientes, tales manipulaciones constituirían *ipso facto* una manipulación de la realización física. Pero, si este es el caso, para toda atribución causal que relacione una propiedad superveniente con su supuesto efecto, existiría otra atribución causal que relaciona su realización física con el mismo efecto. Ya que tenemos buenas razones para aceptar el cierre causal de la física y rechazar la sobredeterminación sistemática de los fenómenos físicos, entonces deberíamos dar la prioridad a la atribución causal concerniente a las realizaciones físicas. De esa manera volvemos a la conclusión de Kim. De hecho, podemos considerar esa segunda objeción de Baumgartner a los evidencialistas como la versión intervencionista del problema de la exclusión causal.

Woodward (2015, 2018) ha intentado solucionar el primer problema modificando su definición original de «manipulación posible» de manera que incluya, como posibles, las manipulaciones de propiedades supervenientes. Aún admitiendo que tal definición modificada sea satisfactoria, eso no ayudaría, sin embargo, con la segunda objeción de Baumgartner. Responder al argumento de Baumgartner no requiere simplemente mostrar que las propiedades

supervenientes pueden ser manipuladas, sino también que, de alguna manera, tales manipulaciones son diferentes de las manipulaciones de las propiedades físicas de las realizan.

Algunos filósofos (por ejemplo, Raatikainen 2010; véase también List & Menzies 2009) han argumentado que las manipulaciones de las propiedades supervenientes apoyan atribuciones causales en las cuales las causas son *proporcionales* a sus efectos (Yablo 1992), mientras que las manipulaciones de los realizadores físicos no. El problema de esa propuesta es que parece asumir, en primer lugar, que el debate es exclusivamente sobre el *papel explicativo* de las atribuciones causales y, en segundo lugar, que las explicaciones basadas en relaciones causales entre propiedades supervenientes son más informativas que aquellas basadas en relaciones causales entre realizaciones física. Pero ese no es el caso: frecuentemente, lo que buscamos son explicaciones reduccionistas, que se refieren a relaciones causales entre realizadores (Baumgartner 2017, p. 13).

En este artículo, sondaré otra estrategia para enfrentar la versión intervencionista del argumento de la exclusión causal. De manera general, esa estrategia consiste en rechazar una de las premisas de Baumgartner, esto es, la premisa según la cual la manipulación de propiedades supervenientes comporta *ipso facto* la manipulación de sus realizadores físicos. Esta premisa se basa en el presupuesto implícito de que, para ser posible, una manipulación tiene que ser, en línea de principio, *físicamente* posible. Sin embargo, Woodward ha repetidamente insistido (Woodward 2003, 2016) sobre el hecho de que, para soportar atribuciones causales verdaderas, las manipulaciones sólo necesitan ser *teóricamente* posibles. Mi tesis es que, si Woodward está en lo cierto, entonces es teóricamente posible manipular propiedades supervenientes independientemente de sus realizaciones físicas.

Para soportar esa tesis, procederé de la siguiente manera. Dado que la concepción intervencionista de la causación es ampliamente derivada del trabajo de Spirtes et al. (2001) y Pearl's (2009) sobre modelos causales, introduciré, en la sección 2, la terminología necesaria para entender la relación entre modelos causales y manipulación. En la sección 3, mostraré como el problema de la exclusión causal puede ser representado (y, de hecho, es representado, por parte de muchos evidencialistas y Baumgartner) por medio de modelos causales modificados, que yo llamaré *modelos causales multi-nivel*. En la sección 4, discutiré el intento de Woodward de modificar la noción de «manipulación posible» para acomodarla a los modelos causales multi-nivel, y así responder a la primera objeción de Baumgartner (2013, 2017). En las secciones 5 y 6, esbozaré mi estrategia para rescatar la causación superveniente en contra del desafío

epifenomenalista. Finalmente, consideraré posibles objeciones a mi perspectiva y expondré alguna consideración general sobre la relación entre intervencionismo y metafísica.

§2. Modelos casuales, intervenciones y causación

Los modelos causales son modelos matemáticos elaborados con la intención de representar y/o descubrir relaciones causales entre variables en un sistema (Hitchcock 2018). Las variables consideradas en un modelo causal son funciones que atribuyen valores numéricos a la ocurrencia o no-ocurrencia de un evento, los estados posibles de un proceso, la existencia o no-existencia de un objeto o, más comúnmente, la posesión de una propiedad (discreta o continua) por parte de un individuo o una población. Cada variable en un modelo causal debe poder tomar por lo menos dos valores. En un modelo causal, dos variables X y Y están causalmente relacionadas *si y solamente si* el valor de Y covaría con el valor de X *como consecuencia* de una manipulación de X.

Formalmente, las relaciones causales entre variables en un sistema pueden ser representadas en una variedad de maneras. Las ecuaciones estructurales y los grafos causales son ciertamente las más comunes. Las ecuaciones estructurales tienen la ventaja sobre los grafos causales de que ofrecen información cuantitativa sobre la relación causal entre variables. Sin embargo, los grafos causales ofrecen —debido a sus características diagramáticas— una representación más inmediata de la estructura causal del sistema bajo estudio. Dado que en este contexto no estamos interesados en cualquier análisis cuantitativo específico, representaré modelos causales exclusivamente por medio de grafos causales.

En la figura 1, es posible apreciar un modelo causal bastante sofisticado, elaborado por el bioestadístico Andrew Forbes y sus colaboradores, con el objetivo de representar la relación causal entre el tabaquismo (X) y el asma (Y). Las otras variables representan un conjunto de factores que interactúan con estas dos variables, tales como la presencia de asma en los padres (D), la presencia de bronquitis crónica (E), o el género (F).

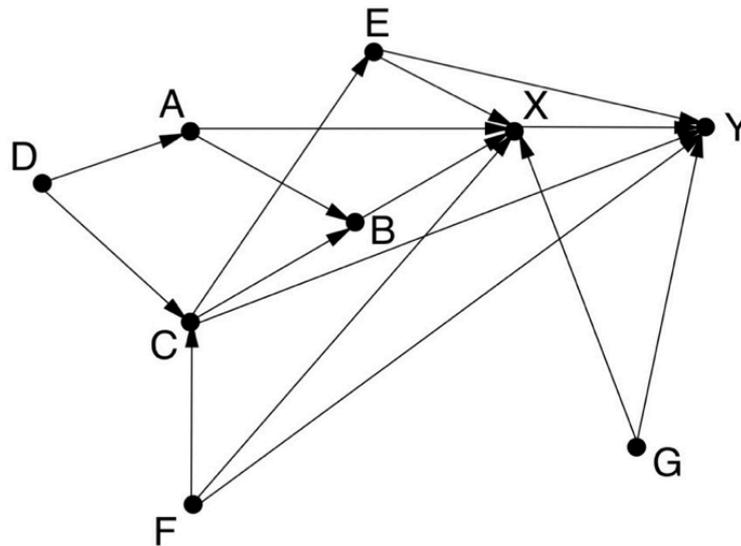


Figura 1: Modelo causal de Forbes et al. de la relación entre tabaquismo (X) y asma (Y) (in Pearl 2018, p. 164)

Cada flecha denota una relación causal, esto es —como acabamos de ver— una relación de covarianza entre la variable–efecto y la variable–causa *debida* al cambio, inducido por manipulación, en el valor de la variable–causa. Esa relación es también llamada de relación de *invarianza* (Woodward 2003). Las relaciones invariantes son tomadas como relevantes para evaluar una clase de enunciados contrafactuales, tales como «Si X tuviese otro valor, Y también tendría otro valor» (por ejemplo, en el presente contexto: «Si un individuo fumase menos, no tendría problemas de asma»), por un cierto conjunto de valores de X. Debido a tal característica de las relaciones invariantes, se dice frecuentemente que, en un modelo causal, una variable es causalmente eficiente si funciona como *difference-maker* por una o más variables.

Cuando un científico empieza a estudiar un nuevo dominio de fenómenos, su situación, en lugar de ser la que es representada en la figura 1, es más parecida a la que se puede observar en la figura 2.

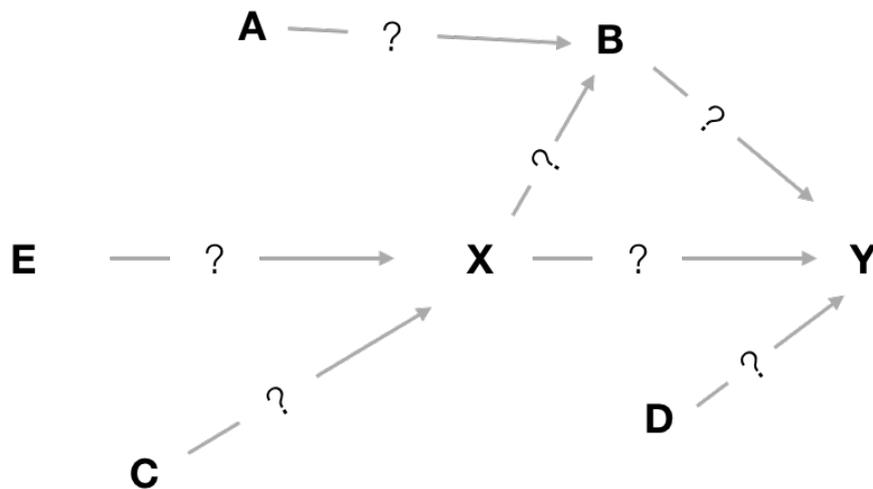


Figura 2: Un conjunto de variables antes de un análisis causal. Las variables conectadas con flechas marcadas con puntos de interrogación están correlacionadas, pero sin ulterior información es imposible decir si están causalmente relacionadas.

Antes de un análisis causal, los investigadores ya poseen, normalmente, informaciones relacionadas con la covarianza entre variables (este hecho es representado en la figura 2 por medio de las flechas marcadas con puntos de interrogación). Sin embargo, sin análisis causal, ellos no pueden decir si dos variables, por ejemplo B y Y, están causalmente relacionadas o son ambas efectos de una causa común, X. Para poder excluir una mera correlación, tenemos que *intervenir* sobre la supuesta causa y ver si esa intervención produce algún cambio también en el supuesto efecto. La posibilidad de manipular el sistema es lo que permite distinguir entre correlación y causalidad. La covarianza entre variantes meramente correlacionadas queda interrumpida (*disrupted*) por las manipulaciones relevantes.

Pero ¿Qué es exactamente una intervención/manipulación? Woodward (2003) la caracteriza de la siguiente manera. Dadas dos variables X y Y, I vale como una manipulación de X con respecto a Y si y solamente si:

(IV)

1. I causa X;
2. I actúa como un interruptor para todas las otras variables que causan X. Esto es, ciertos valores de I son tales que, cuando I alcanza tales valores, X deja de depender del valor de otras variables que causan X y depende únicamente de los valores de I;
3. Cualquier ruta directa entre I y Y pasa por X. Esto es, I no causa directamente Y y no es un causa de cualquier causa de Y que es distinta

de X excepto, claramente, por aquellas causas de Y, si es que existen, que están construidas en la conexión $I \rightarrow X \rightarrow Y$ misma; esto es, excepto por (a) cualquier causa de Y que es efecto de X (o sea, una variable causal que se encuentra entre X y Y) y (b) cualquier causa de Y que está entre I y X y que no tiene efectos sobre Y independientemente de X;

4. I es (estadísticamente) independiente de cualquier variable Z que causa Y y que está en una ruta directa que no pasa por X (Woodward 2003, p. 98).

Como consecuencia de 2, la historia causal de X es completamente excluida (*screened off*) por I. 3 y 4 expresan el hecho de que I requiere que cualquier variable externa a la ruta (correspondiente a una de las «flechitas» en las figuras arriba) entre X y Y sea *mantenida fija*. Una intervención que satisfaga tales criterios es considerada una intervención «quirúrgica» (Woodward 2016), dado que aísla completamente la relación entre X y Y de cualquier posible factor perturbador. Las intervenciones quirúrgicas definen el test ideal para determinar si la relación entre X y Y es realmente invariante (y por lo tanto causal) o no.¹

Es fundamental tener en mente que, para los intervencionistas, la posibilidad de una intervención quirúrgica exitosa sobre una variable no sólo es una condición suficiente para establecer la relevancia causal de la variable, sino que es también una condición necesaria. Una manipulación exitosa de una variable con respecto a otra (esto es, una manipulación en la que la segunda variable cambia como consecuencia de la intervención sobre la primera variable) permite automáticamente atribuir una relación causal entre ellas. Inversamente, sin embargo, si no somos capaces de realizar una intervención quirúrgica sobre una variable, eso no significa simplemente que no podemos testar su papel causal, sino que tal variable *no tiene* un papel causal (por lo menos con relación a su supuesta variable–efecto).²

¹ Claramente la noción de intervención es ella misma una noción causal. Eso no debe ser tomado como una limitación de la teoría de Woodward (2003), ya que su finalidad no es la de ofrecer un análisis «reductivo» de la noción de causa en términos de nociones no causales (como es el caso, por ejemplo, de Lewis 1986 y de muchos otros autores). Más bien, lo que Woodward persigue es una caracterización rigurosa de lo que permite a los científicos decir que una variable *específica* es causalmente eficaz, sin pretensiones de agotar la variedad de connotaciones de la noción de causa.

² En la literatura, el hecho de que la posibilidad de una intervención quirúrgica sea condición necesaria y suficiente para la causación es usualmente establecida más rigurosamente por la que Woodward llama *condición (M)* (Woodward 2003, p. 59). He evitado citarla explícitamente aquí porque contiene una distinción entre causas *directas* y *contribuidoras* que es irrelevante en el presente contexto. Además, la

§3. La caracterización intervencionista del problema de la exclusión causal

Una restricción común a la elección de variables en los modelos causales es que los modelos causales deben satisfacer la llamada *condición causal de Markov*, la cual establece que una variable debe ser independiente de cualquier otra variable excepto de sus efectos (Hausman & Woodward 1999; Spirtes et al. 2001; Eronen 2012). Entre otras cosas, la condición causal de Markov impide que los modelos causales representen relaciones de dependencia no-causal. Esta consecuencia de la condición causal de Markov es generalmente bienvenida porque las relaciones de dependencia no causal entre variables pueden viciar nuestras inferencias causales.

Spirtes and Scheines (2005) discuten, por ejemplo, el hipotético caso en el cual una cierta enfermedad cardíaca (HD) es aparentemente dependiente de tres variables que representan la cantidad total de colesterol (TC), la cantidad de lípidos de alta densidad (HDL), y la cantidad de lípidos de baja densidad (LDC). Dado que TC se *define* como HDL + LDL, es imposible manipularlo manteniendo HDL y LDC fijos (como requeridos por las cláusulas 3 y 4 arriba): una manipulación de TC *implica* (lógicamente) la manipulación de HDL y LDC. Dado, sin embargo, que la posibilidad de una intervención quirúrgica es una condición necesaria para que una variable pueda ser considerada causalmente eficiente, eso tendría la consecuencia que TC no es causalmente relacionado con HD. Pero eso es contraintuitivo y, en última instancia, absurdo: ¿como es posible que la cantidad de lípidos de baja densidad y los de alta densidad tengan, si tomados individualmente, relevancia causal con relación al desarrollo de una cierta enfermedad cardíaca, pero no la tienen si son tomados en conjunto?

La relación entre TC, por un lado, y HDL y LDL, por el otro, es una relación que puede ser llamada de «dependencia definicional». Los investigadores pueden evitar tales situaciones ambiguas omitiendo la variable definida (en ese caso TC) del modelo. Aunque diferente de la dependencia definicional, la

condición (M) relativiza las atribuciones causales a un conjunto de variables V de una manera que es potencialmente engañosa. La relación entre causalidad en V y causalidad *simpliciter* ha sido aclarada por Woodward (2008) y Baumgartner (2013), y no necesitamos volver a esa cuestión. En este contexto, asumiré simplemente que, aunque las atribuciones causales sean siempre formuladas por investigadores que trabajan con un conjunto específico —y restringido— de variables, eso no significa que las atribuciones sean relativas a tal conjunto. Los modelos causales que representas adecuadamente un cierto dominio de fenómeno no pueden, idealmente, contener información contradictoria, ya que supuestamente representan relaciones causales «reales» (esto es, no dependientes del modelo. Volveré brevemente sobre esta cuestión al final de la sección 6).

relación de superveniencia en cuestión en el problema de la exclusión causal es una relación de dependencia tradicionalmente entendida como no-causal. Por lo tanto, en teoría, no podría ser representada en un modelo causal. No obstante, la mayoría de los autores que participan en el presente debate han explícitamente usado, en alguna ocasión, grafos que violan la condición causal de Markov. Una vez que se aclare que tales grafos son, por así decirlo, «artefactos filosóficos», y no modelos causales en el sentido propio de Spirtes et al. o Pearl, no veo ningún problema en usarlos como recurso argumentativo (contrariamente a lo que ha sido defendido, por ejemplo, por Eronen y Brooks 2014).

Más allá de las idiosincrasias de cada autor, es posible caracterizar genéricamente un modelo causal que incluye relaciones de superveniencia como un modelo causal «multi-nivel». Un modelo causal multi-nivel es un modelo en el cual algunas de las variables covarían con otras por razones *metafísicas* (Woodward 2015, 2018). Más específicamente, la razón metafísica por la cual una variable superveniente (esto es, una variable que representa una propiedad, evento o proceso considerado superveniente) covaría con la variable realizadora (esto es, la variable que representa una propiedad, evento o proceso considerado como la realización física de la propiedad, evento o proceso superveniente) es el cierre causal de la física. De acuerdo con tal principio, no puede haber diferencias en propiedades (o, por extensión, eventos, procesos) estudiadas por las ciencias especiales sin una diferencia en las propiedades estudiadas por la física.

Dada esta caracterización de los modelos causales multi-nivel, estamos listos para formular el problema de la exclusión causal en términos intervencionistas. Imaginemos que queremos saber si B, que es físicamente realizada por A, es causalmente relevante por Y (figura 3):

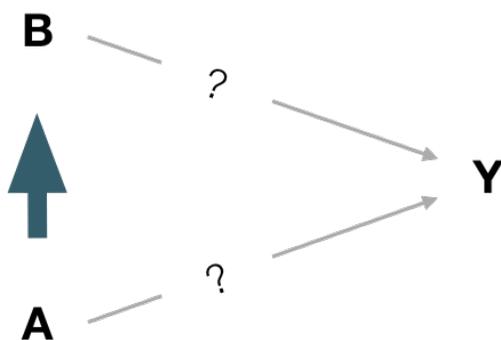


Figura 3: un hipotético modelo causal multi-nivel (antes del análisis causal).

De acuerdo con lo que hemos visto en la sección 2, para poder decir que B es causalmente relevante por Y, necesitamos encontrar una posible manipulación de B que satisfaga los requisitos para ser considerada una intervención quirúrgica y, por supuesto, que tenga como efecto un cambio correspondiente en Y. De acuerdo con Baumgartner (2009), tal intervención es imposible, porque cualquier manipulación de B viola las cláusulas 3 y 4 de (IV).

Con relación a 3, dado que un cambio en B —la propiedad superveniente— implica un cambio en A —la realización física— no es verdad que todas las rutas entre I y Y pasan a través de B. Por el contrario, hay una ruta entre I y Y que pasa por otra variable (A). De manera similar, considerando ahora la cláusula 4, I no es (estadísticamente) independiente de cualquier variable Z que causa Y y que está en la ruta directa que no pasa por B. Por el contrario, I es dependiente de una variable (A) que causa Y y está en una ruta que no pasa por B.

Para simplificar: cualquier cambio en una variable superveniente implica covariancia con la base realizadora, tal que es imposible manipular B sin manipular simultáneamente A. Dado que, como vimos, el hecho de que la intervención sea quirúrgica es una condición suficiente y necesaria para que la relación bajo estudio pueda ser considerada como causal, si no podemos manipular quirúrgicamente B, eso significa que B es (por lo menos con respecto a Y) causalmente inerte, puramente epifenoménico. Esta es, claramente, una conclusión completamente indeseable para los evidencialistas.

§4. Intentos de solución y nuevos problemas

La manera aparentemente más promisoría de replicar a la versión intervencionista del problema de la exclusión causal es argumentar que los modelos causales multi-nivel no funcionan exactamente como los modelos causales normales. Esa estrategia, que ha sido adoptada, de una manera u otra, por la mayoría de los evidencialistas (Shapiro & Sober 2007; Raatikainen 2010; Weslake *forthcoming*; Eronen 2012), ha sido más claramente definida por Woodward (2015, 2018). Por esa razón, me concentraré aquí exclusivamente en su propuesta. Para incluir las manipulaciones de variantes supervenientes en el conjunto de las manipulaciones quirúrgicas posibles, Woodward sugiere la siguiente reformulación de las cláusulas 3 y 4 de (IV):

(IV')

3'. Cualquier ruta directa de I y Y pasa por X o por una variable Z que está relacionada con X en términos de superveniencia.

- 4'. I es (estadísticamente) independiente de cualquier variable Z que causa Y y que está en una ruta directa que no pasa por X, *excepto que por su realización física*.

De esa manera, las manipulaciones quirúrgicas de variantes supervenientes están admitidas y, por lo tanto, es aparentemente recuperada la posibilidad de atribuirles eficacia causal.

Aunque, a primera vista, esta podría parecer una solución *ad hoc* del problema, es fácil ver que tiene un fundamento bastante razonable. Cuando un científico dice que la ingestión de un fármaco causa la recuperación de una enfermedad, no quiere claramente decir que es posible manipular la propiedad curativa del fármaco —la variable superveniente— mientras que su estructura química —la realización física— es mantenida fija (Shapiro & Sober 2007; Woodward 2015). La imposibilidad de intervenir en una variable superveniente mientras que la realización física es mantenida fija no parece, sin embargo, invalidar la atribución causal según la cual es la ingestión del fármaco —*en lugar que* su no-ingestión— lo que causa la recuperación de la enfermedad. Generalizando el punto (fundamental también para entender la propuesta elaborada más adelante, en las secciones 5 y 6): la imposibilidad de manipular las propiedades supervenientes mientras que se mantienen fijas las realizaciones físicas no parece impedir que los científicos avancen hipótesis causales sobre propiedades supervenientes. Woodward (así como otros evidencialistas) toma las prácticas científicas como más relevantes para la comprensión de nuestras atribuciones causales que cualquier argumento metafísico. Si los científicos consideran que una determinada propiedad superveniente es causalmente eficaz, los filósofos tienen que adaptar su noción de causalidad a tal práctica, y no insistir en la aparente solidez del argumento de la exclusión causal.

Aunque de alguna manera favorable a la justificación que Woodward ofrece por la reformulación de (IV), Baumgartner (2013, 2017) niega que tal reformulación sea suficiente para evitar la amenaza del epifenomenalismo. Mientras que con (IV) el problema era que ese criterio no permitía atribuir eficacia causal a las variables supervenientes, con (IV') el problema es que no permite distinguir las atribuciones que se refieren a las variables supervenientes de aquellas que se refieren a sus realizaciones físicas. Dado que las atribuciones causales que se refieren a variables supervenientes no añaden ninguna información causal a aquellas que relacionan sus realizaciones físicas, Baumgartner piensa que, por razones de simplicidad, deberíamos preferir las segundas a las primeras. Más en detalle, el argumento de Baumgartner puede ser

reconstruido de la siguiente manera:

- P1. Debido al cierre causal de la física, cualquier manipulación de una variable superveniente, B, es *ipso facto* una manipulación de su realización física, A;
- P2. Para cualquier manipulación exitosa de B con respecto a Y existe otra manipulación de A —*pero no de B*— igualmente exitosa;
- P3. Manipulaciones de A que no son manipulaciones de B son «más quirúrgicas» que las manipulaciones de B;
- P4. En ausencia de otras razones, las hipótesis causales obtenidas a partir de intervenciones «más quirúrgicas» son preferibles a las hipótesis causales obtenidas a partir de intervenciones «menos quirúrgicas»;
- C. Las hipótesis causales obtenidas a partir de manipulaciones de realizadores físicos son preferibles a las hipótesis obtenidas a partir de manipulaciones de variables supervenientes.

Las manipulaciones de variantes supervenientes son, para Baumgarten, «torpes» (*fat-handed*; Scheines 2005). Una manipulación torpe es una intervención que funciona como causa común de dos (o más) variables. Es por esa razón que las manipulaciones de variables supervenientes son «menos quirúrgicas» que las manipulaciones realizadas únicamente sobre las realizaciones físicas correspondientes. Nótese que la conclusión de Baumgartner no es tan pesimista como la que se deriva del argumento de la exclusión causal originario, de Kim. Baumgartner no dice que la causación superveniente es imposible sino que, más modestamente, el intervencionismo, si no asume ulteriores premisas, no es capaz de dar cuenta de esa.

Como he mencionado en la introducción, algunos autores (e.g., Raatikainen 2010; List & Menzies 2009; inspirados por Yablo 1992) parecen ver la cuestión de manera diferente de Baumgartner. Desde el punto de vista de estos autores, las atribuciones causales ofrecidas en términos de variables supervenientes son más informativas que las atribuciones causales ofrecidas en términos de realizadores físicos porque, en las primeras, las supuestas causas son más *proporcionales* a sus efectos. Aunque un análisis detallado de la noción de proporcionalidad no puede ser ofrecido aquí, podemos decir que una causa C1 es más proporcional a un cierto efecto E que una causa alternativa C2 si y solamente si E es contrafactualmente dependiente de C1 en más mundos posibles de aquellos en los que es contrafactualmente dependiente de C2 (asumiendo una versión

levemente modificada de la semántica de mundos posibles Lewis–Stalnaker). Dado que las variables supervenientes son múltiplemente realizables, hay mundos en que una diferencia en la realización física A instancia la misma variable superveniente B pero no produce ninguna diferencia en E. Diferencias en B implican diferencias en A que producen diferencias en E en más mundos posibles. Por esta razón, aún admitiendo el cierre causal de la física, citar B como causa de E es más «específico» e informativo que citar A.

Estoy de acuerdo con Baumgartner (2017) en que apelar a la proporcionalidad no ayuda a disolver completamente el escepticismo respecto a la capacidad del intervencionismo de atribuir relevancia causal a las variables supervenientes. Los proporcionalistas asumen que el hecho de que las atribuciones causales ofrecidas en términos de variantes supervenientes sean contrafactualmente verdaderas en más circunstancias que las atribuciones causales ofrecidas en términos de realizaciones físicas es una evidencia de que las variantes supervenientes son causalmente eficaces. No obstante, podría ser el caso que las variantes supervenientes sean simplemente buenos «indicadores» (*proxies*) de causalidad. Lo que permite que los científicos que emplean modelos causales distingan causación y correlación es la posibilidad de manipular quirúrgicamente las supuestas causas. Pero si (como sugiere el argumento de Baumgartner) tales manipulaciones quirúrgicas no pueden ser realizadas sobre las variables supervenientes, no podemos decir si la dependencia contrafactual, por parte del efecto bajo estudio, de estas variables es genuina o un mero artefacto resultante del modelo. Nadie duda de que las explicaciones que citan variables supervenientes pueden ser buenas explicaciones, pero eso no garante *ipso facto* que las variantes supervenientes sean eficaces causalmente.

Otro problema con la proporcionalidad (posiblemente debido al hecho de que los proporcionalistas, así como la mayoría de los autores involucrados en el debate, se concentra casi exclusivamente en la causación mental) es que conduce a consecuencias que son claramente falsas. Simplemente no es verdad que, en todos los casos, la mejor explicación causal posible es la que es ofrecida en términos de variables proporcionales. En muchos casos —como cuando intentamos ofrecer una explicación mecanicista (Craver 2007)— nos concentramos explícitamente en las características de «bajo nivel» de un fenómeno para explicar las características de «alto nivel». Aunque estas explicaciones no satisfacen el criterio de proporcionalidad, frecuentemente satisfacen otros criterios igualmente importantes (parsimonia, generalizabilidad, etc.). La proporcionalidad es solamente un criterio entre otros, en la elección de variables. En algunos contextos quizás sea el más importante, pero no puede ser considerado como una guía infalible para detectar la eficacia causal de una

variable.

§5. Una estrategia alternativa

Después de haber presentado críticamente el estado actual del debate, estoy finalmente listo para presentar mi propuesta. A lo largo de esta y de la próxima sección esbozaré el siguiente argumento. En primer lugar, discutiré un aparente caso de manipulación torpe en un modelo causal tradicional (no multi-nivel) y mostraré que, en realidad, conduce a una atribución causal perfectamente aceptable. En segundo lugar, mostraré que la razón por la cual tal atribución es aceptable es que, de acuerdo con la concepción de Woodward (2003, 2016), una manipulación «posible» sólo necesita ser *teóricamente*, y no físicamente, posible. En la próxima sección, concluiré que esta concepción permite, por lo menos en teoría, atribuir eficacia causal a variables supervenientes.

Pasemos, entonces, a la primera parte del argumento. Tómese la siguiente afirmación:

- (G) La atracción gravitacional de la Luna causa el movimiento de las mareas.

Dado que cualquier intervención que altere la posición de la Luna tendrá, muy probablemente, un impacto independiente sobre las mareas (porque alterará, por ejemplo, la posición de algún otro objeto masivo que, a su vez, tiene algún efecto sobre las mareas), podría argumentarse que (G) no es una atribución causal verdadera, porque no satisface (IV) (viola la cláusula 3). Alternativamente, se podría decir que (G) se basa en una manipulación torpe, ya que, para que sea verdadera, requiere que se manipule la posición de la luna *junto con* aquella de otros objetos masivos.

Por supuesto, la idea de que no es *posible* manipular la posición de la luna sin alterar la posición de otros objetos masivos se basa en que no sería *físicamente posible* realizar tal intervención.³ Sin embargo, Woodward (2016) considera que no es ese el sentido en el que tenemos que entender la noción de manipulación

³ Nótese, *en passant*, que ninguno de los participantes en el debate requiere que la posibilidad física de una intervención sea una posibilidad *de facto*. Claramente no podemos, en la práctica, cambiar la posición de la luna (así como no podemos, en la práctica, alterar fenómenos del pasado). Sin embargo, cuando hablamos de manipulaciones físicamente posibles, estamos pensando en alteraciones (quirúrgicas) que podrían, *en un principio*, ser realizadas físicamente.

posible. (G) es una atribución causal supuestamente verdadera porque:

Tenemos una base de principio en la mecánica Newtoniana y en la teoría gravitacional misma para contestar preguntas con respecto a lo que ocurriría si tales intervenciones quirúrgicas [i.e., intervenciones capaces de alterar la posición de la luna sin alterar la de otros planetas] ocurrieran y eso es suficiente para justificar la atribución causal (G) (Woodward 2016, §10, sin página).

Woodward sigue argumentando de la siguiente manera:

Desde este punto de vista, lo que es crucial no es si el antecedente del contrafactual relevante es nomológicamente o contrafactualmente posible sino si poseemos teorías científicas bien fundadas y una matemática concomitante que nos permite contestar de manera fiable preguntas sobre lo que ocurriría bajo la suposición de tales antecedentes (Woodward 2016, §10, sin página).

En otras palabras, lo que permite asegurar que (G) cumple las condiciones previstas por (IV), y que por lo tanto es una atribución causal supuestamente verdadera, es que poseemos una teoría que nos dice que pasaría *si* tal manipulación fuera físicamente posible y no torpe.

Nótese que decir que, para satisfacer (IV), una manipulación debe ser teóricamente posible no es equivalente a decir que debe ser sólo *lógicamente* posible, como algunos autores (Reutlinger 2013; Gebharder 2017) parecen pensar. Ciertamente, manipulaciones lógicamente o conceptualmente inconsistentes no pueden soportar ninguna atribución causal. Sin embargo, una manipulación teóricamente posible tiene que ser también *bien definida*, en el sentido que debe ser claro cuales aspectos de la (o las) teoría(s) empleadas para testar el contrafactual en cuestión son llamados a apoyarlo (Woodward 2003, p. 115).

Tómese, por ejemplo, la siguiente —ciertamente bizarra— afirmación (un caso análogo es discutido por Woodward 2003, 2016):

(C) El hecho de que Tom sea un gato causa que tome leche.

Para que (C) pueda ser considerada una atribución causal verdadera es necesario que algo como el siguiente contrafactual sea verdadero: «Si Tom fuera un lagarto, entonces no tomaría leche». Para que ese contrafactual sea verdadero, debe ser posible intervenir sobre Tom de manera que deje de ser un gato y pase a ser un

lagarto. Tal manipulación no parece ser lógicamente o conceptualmente imposible; sin embargo, no queda para nada claro lo que deberíamos hacer para que Tom cambiara de especie. Esto es, aunque no lógicamente imposible, la manipulación no es bien definida. Y es por eso que (C) no es una atribución causal aceptable.

Confróntese ahora (C) con la siguiente atribución causal (derivada de Holland 1986 y Woodward 2003, p. 115 ss.):

(D) Ser mujer es causa de discriminación laboral y salarial.

En este caso, además de ser lógicamente posible, es también físicamente posible manipular el sexo de una persona. Sin embargo, y este es el aspecto crucial de la discusión, no parece que la posibilidad física sea relevante para determinar si (D) es verdadero o falso.

Lo que se entiende por (D) es, plausiblemente, algo como:

Los empleadores determinan si contratar A y cuanto pagar A por lo menos en parte con base en sus creencias sobre el género de A, independientemente de lo que piensen sobre las cualificaciones, credenciales, formación etc. de A; esto es, diferencias en las creencias del empleador con relación al género de los postulantes causan diferencias en si los postulantes recibirán una oferta de trabajo o en el sueldo que les será ofrecido (Woodward 2003, p. 115).

Lo que es razonable manipular para dar cuenta de (D) son un conjunto de prácticas laborales y de creencias personales.⁴ Estas difícilmente pueden ser manipuladas en un sentido «físicamente» quirúrgico y no torpe (intervenir en un aspecto específico de tales prácticas llevaría, con toda probabilidad, a cambiar muchas otras condiciones de trasfondo). Sin embargo, eso no parece importar con respecto a la aceptación de (D) como una genuina atribución causal. Lo que lleva, en este caso, estadísticos y científicos sociales a aceptar (D) es que existen modelos y teorías de referencias (y herramientas matemáticas) que permiten evaluar empíricamente su valor de verdad por medio de experimentos u observaciones controladas.

⁴ Woodward, por ejemplo, sugiere un experimento en el que los *curricula* de los postulantes son entregados al empleador con los datos relacionados al género invertidos (las mujeres aparecen con un nombre masculino ficticio, y vice versa), para determinar si eso influye en las probabilidades de obtener un trabajo.

§6. Manipulando sombras: un caso de estudio y alguna sugerencia para la disolución del problema de la exclusión causal

Vimos en la última sección que, en línea de principio, lo que hace que una atribución causal sea considerada legítima es que el antecedente del contrafactual que la apoya sea *teóricamente* manipulable. Esto es, no es necesario que las manipulaciones que apoyan atribuciones causales sean *físicamente* quirúrgicas, sino solamente que sean claramente definidas de acuerdo con alguna teoría o modelo de referencia. En el caso de los modelos causales multi-nivel, las cosas son aparentemente más complicadas debido al hecho de que, supuestamente, la manipulabilidad de las variables supervenientes *depende* de la manipulabilidad de los realizadores físicos. Como vimos en secciones anteriores, la razón de eso es que, de acuerdo con el principio del cierre causal de la física, las variables supervenientes no pueden cambiar sin que haya un cambio en los realizadores físicos.

Mi tesis es que el hecho de que una variable superveniente covaríe, por razones metafísicas, con una realización física no implica que la primera sea causalmente inerte, ya que es posible que la manipulación de la variante superveniente sea independiente de la manipulación de la realización física, en virtud de la *caracterización teórica* de tal manipulación. Esto es, por razones análogas a las que vimos en la sección anterior, podemos atribuir eficacia causal a una variable superveniente aunque no sea físicamente posible manipularla de manera independiente de sus realizaciones físicas si existe una teoría o un modelo que definan claramente en que consistiría tal manipulación.

Ilustraré esa tesis con un ejemplo, derivado de Weber (2008). En ese artículo, el autor discute un importante modelo en la historia de las neurociencias, realizado por A. L. Hodgkin y A. F. Huxley (1952). El principal resultado de estos dos fisiólogos fue el de ofrecer la que es conocida como ecuación de corriente total, la cual define la corriente total entre membranas neuronales como función de la capacidad de la membrana, el voltaje, el equilibrio potencial del sodio y del potasio y la conductancia máxima del sodio y del potasio, entre otras variables. El modelo, del punto de vista de Weber, puede ser dividido en dos partes: por un lado, contiene un modelo de *conductancia*, que describe como la permeabilidad de la membrana depende del voltaje; por el otro, contiene un modelo de *acción*, que describe como estos cambios en la permeabilidad son realizados por diferencias en el potencial electroquímico entre las membranas neuronales.

De acuerdo con nuestra terminología, podemos decir que los fenómenos descritos en el modelo de acción constituyen la realización física de los fenómenos descritos en el modelo de la conductancia. Podría, por lo tanto,

construirse un modelo causal multi-nivel (extremadamente simplificado) con tres variables —(V) voltaje; (E) potencial electroquímico entre membranas neuronales; y (P) permeabilidad de la membrana— tal como en la figura 4.

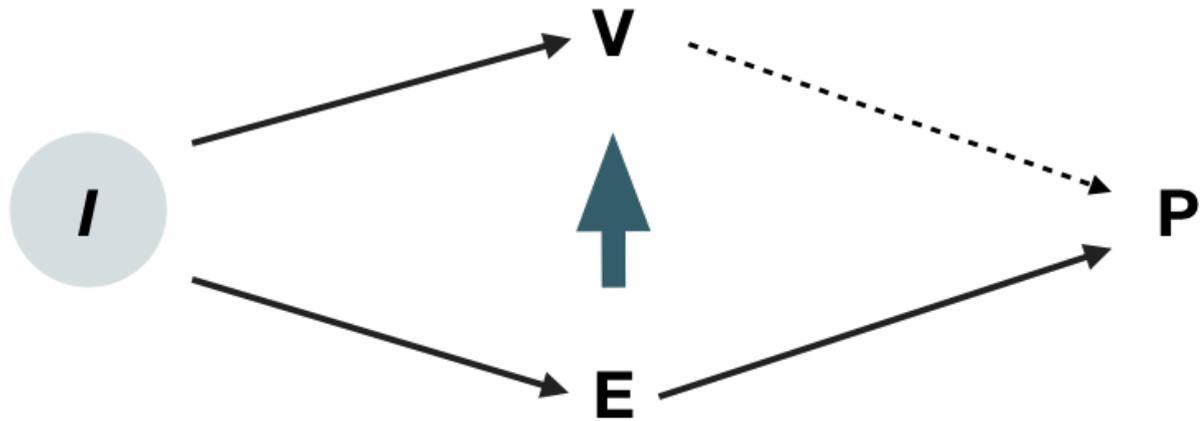


Figura 4: un modelo causal multi-nivel simplificado del modelo de Hodgkin y Huxley. La línea discontinua es la que, según los sostenedores de la exclusión causal, debería quedar excluida.

Para el defensor del argumento de la exclusión causal, el hecho de que las diferencias en voltaje sean realizadas por diferencias en el potencial electroquímico vuelve el papel causal de V respecto a P epifenoménico.

Sin embargo, considérense los siguientes aspectos de la cuestión, ambos señalados por Weber. En primero lugar, resulta que el modelo de acción de Hodgkin y Huxley es equivocado; esto es, los factores que los dos fisiólogos identificaron como determinantes para la transferencia de iones entre membranas eran incorrectos. Sin embargo, aún hoy en día se considera el modelo de conductancia como substancialmente correcto. Es más, es justamente el hecho de que el modelo haya sido confirmado experimentalmente lo que ha permitido, en las décadas siguientes, descubrir la base física correcta del fenómeno de permeabilidad de la membrana (Weber 2008, p. 1001). En segundo lugar, nótese que Hodgkin y Huxley se refieren con un vocabulario causalista a la relación causal entre V y P, *aún cuando* reconocen que tal relación causal debe ser realizada físicamente.

La razón por la que Hodgkin y Huxley adoptan una interpretación causalista, y no epifenoménica, del modelo de conductancia es —siguiendo Weber— que el modelo es formulado en términos de variables (en nuestro modelo simplificado, el voltaje entre membranas) que pueden ser manipuladas. La manipulación de tales variables —este es el aspecto central de la cuestión— es

posible en virtud del papel que ellas juegan en el modelo, y es (teóricamente) *independiente* de las condiciones que permiten manipular la base física. De hecho, Hodgkin y Huxley no conocían la base física del proceso y, por lo tanto, no podían realizar manipulaciones quirúrgicas en ella.

Claramente, el defensor de la exclusión causal podría objetar que, aunque Hodgkin y Huxley no conocían la base física del proceso, para poder manipular las variables supervenientes ellos *tenían que* manipular alguna base física. Pero es precisamente aquí que la intuición del epifenomenalista falla en apreciar la perspectiva del evidencialista. Como ya dijimos en la sección 4, para el evidencialista la práctica científica tiene prioridad conceptual sobre cualquier presupuesto metafísico. Aún cuando, en virtud del principio del cierre causal de la física, nos parecería razonable inferir una relación causal más básica respecto a la que se da entre variables reconocidas como supervenientes, esta no debería ser una buena razón para excluir relaciones causales corroboradas entre variables supervenientes. Por el contrario, si la atribución causal es informativa con respecto a la relación entre dos variables (como parece ser en el caso del modelo de conductancia), esta sí parece ser una razón suficiente —para el científico y, por lo tanto, debería serlo también para el filósofo— para no excluirla.

Por supuesto, es un importante *desideratum* de la actividad científica el de descubrir realizadores físicos de propiedades supervenientes. El mismo caso de Hodgkin y Huxley lo revela: aunque ellos fallaron en identificar los realizadores físicos de la transferencia de corriente entre membranas neuronales, otros científicos posteriores siguieron en la búsqueda, y hoy en día tenemos un modelo substancialmente correcto. Por lo que dice Weber, eso no impide que, aún hoy en día, los fisiólogos consideren el modelo de conductancia como un modelo causal.

Esa tensión se puede resolver interpretando el principio del cierre causal de la física como un principio heurístico, más que como una real restricción sobre las atribuciones causales realizadas por los científicos. Esa idea —ya, de alguna manera, relativamente difusa en filosofía de la ciencia (cf. Sarkar 2005)— permite una caracterización más realista de la cuestión de si una variable debería o no ser causalmente eliminada. En lugar que perseguir una respuesta general, aplicable a cualquier escenario, el filósofo debería preguntarse, en cada caso: ¿Como es definida la manipulación de la variable bajo estudio? Si la manipulación de la variable superveniente es definida en términos similares a aquellos en los que es definida la manipulación de su realización física, entonces tenemos —en un principio— buenas razones para adoptar una postura reduccionista y excluir la variable superveniente. Si, por el contrario, la respuesta

a la pregunta es afirmativa, entonces deberíamos aceptar ambas atribuciones causales como genuinas.

Claramente, este criterio de «similitud» entre manipulaciones debe ser caracterizado de manera más rigurosa. Sin embargo, en este artículo, mi objetivo es, más modestamente, el de mostrar que el enfoque evidencialista tiene recursos para enfrentarse al problema de la exclusión causal de una manera más *local* y menos global, considerando caso por caso cuando es oportuno aceptar atribuciones causales sobre variables superveniente, y cuando no.

La presente propuesta tiene algunas importantes similitudes con aquella de Raatikainen (2010) y otros «proporcionalistas», presentada en la sección 4, en el sentido de que parte de consideraciones de tipo epistémico–teórico para resolver un problema considerado tradicionalmente como genuinamente metafísico. Las principales diferencias entre mi propuesta y la de los proporcionalistas es que mi propuesta no reduce el problema de identificar relaciones causales al problema de ofrecer explicaciones causales y, además, no considera la proporcionalidad como criterio para identificar explicaciones causales apropiadas. Sin embargo, al igual que la propuesta proporcionalista, también la presente propuesta dejará insatisfechos a los autores con inclinaciones más metafísicas, ya que hace depender la verdad de las atribuciones causales de nuestras habilidades epistémico–teóricas de manipulación de variables. Al fin y al cabo, dado que somos nosotros los que construimos las teorías que nos permiten manipular variables, las atribuciones causales —podría decirse— no identificarían una referencia objetiva.

Si es interpretada en un sentido estricto, esa objeción es exagerada. Aún sin adoptar una perspectiva realista respecto a las representaciones científicas, cabe esperar que el progreso científico nos permita descubrir relaciones causales que, de alguna manera, son «reales» (aunque tal vez sólo en el sentido de que pueden ser distinguidas de meras correlaciones). Por otro lado, si lo que se quiere decir es que la presente propuesta de solución al problema de la exclusión causal deja en abierto si hay algún *truth-maker* objetivo, más allá de nuestras atribuciones causales, que determina inequívocamente la estructura causal del mundo, entonces es exactamente ese el caso. La perspectiva evidencialista, como ya he discutido arriba, intenta reconstruir una imagen coherente del mundo a partir de lo que los científicos concluyen sobre fenómenos empíricos. Cualquier presupuesto transempírico, como aquel de acuerdo con el que la relación de superveniencia implica exclusión de las causas supervenientes, debería simplemente ser considerado como menos informativo respecto a los conocimientos que derivamos de las prácticas científicas. Si los científicos se

refieren en términos causales a ciertos fenómenos supervenientes, y esas atribuciones son soportadas por la teoría manipulacionista (en los términos precisados en estas últimas dos secciones), eso debería tener prioridad sobre cualquier argumento metafísico.

§7. Conclusión

El problema de la exclusión causal es tradicionalmente tratado como un problema relacionado con la causación mental. En realidad, una versión extendida de ese argumento tiene, potencialmente, consecuencias desastrosas para cualquier atribución causal en las ciencias especiales. Dado que las propiedades y fenómenos estudiados por las ciencias especiales necesitan, de alguna manera, ser realizados físicamente, si el argumento de la exclusión causal es correcto, entonces ninguna atribución causal en biología o sociología sería, estrictamente hablando, verdadera.

En este artículo he considerado como un grupo de filósofos, en ocasiones llamados «evidencialistas», ha intentado solucionar el problema de la exclusión causal gracias a la teoría manipulacionista de la causación (Woodward 2003). De acuerdo con esta teoría, una relación entre dos variables es causal si es posible manipular la supuesta causa de manera a producir cambios correspondientes en el supuesto efecto. Aunque *prima facie* exitosas, las propuestas evidencialistas tienen —como es señalado por Baumgartner (2009, 2010, 2013, 2017)— algunos límites. Después de haber discutido (a lo largo de las secciones 2–4) como el problema de la exclusión causal se configura en una perspectiva manipulacionista, he intentado caracterizar (en las secciones 5 y 6) una posible respuesta a las críticas levantadas por Baumgartner.

Básicamente, he mostrado que si por «manipulación posible» se entiende una manipulación *teóricamente* posible (como, de hecho, se puede extrapolar de Woodward 2003), entonces las variables supervenientes son, por lo menos en algunos casos, manipulables. Esto es: aunque, de hecho, las variables supervenientes requieran realizadores físicos y no puedan, físicamente, ser manipuladas sin manipular sus realizadores físicos, ese sentido *físico* de «manipulación posible» no es el que es relevante para las atribuciones causales en las ciencias especiales. La posibilidad de una manipulación no es determinada, en una perspectiva evidencialista, por el presupuesto metafísico del cierre causal de la física, sino que por lo que los modelos específicos definen como manipulable.

Esta propuesta, aunque necesita ser complementada por una discusión más detallada de *como* las teorías o los modelos científicos caracterizan ciertas

variables como manipulables, pone, de mi punto de vista, las bases para una comprensión de cuestiones tradicionalmente consideradas como metafísicas en una perspectiva más cercana a las prácticas científicas. Si bien esa perspectiva no permite alcanzar grandes conclusiones sobre la existencia de regularidades últimas concernientes la estructura causal del mundo, sí nos acerca a un entendimiento más adecuado de como los científicos —los cuales, a su vez, más se acercan a descubrir tales regularidades— trabajan.*

* AGRADECIMIENTOS: Agradezco el Fondo Nacional de Desarrollo Científico y Tecnológico de Chile (Beca N° 1171017 FONDECYT REGULAR), la Fundação para a Ciência e a Tecnologia de Portugal (FCT – Contrato N° DL57/2016/CP1479/CT0064) y el Conselho Nacional de Desenvolvimento Científico e Tecnológico de Brasil (CNPq – beca n° 402619/2016–1) por el apoyo ofrecido en la realización del presente trabajo de investigación.

REFERENCIAS

- BAUMGARTNER, Michael (2009). «Interventionist Causal Exclusion and Non-Reductive Physicalism». *International Studies in the Philosophy of Science* 23: pp. 161–178. DOI: 10.1080/02698590903006909.
- BAUMGARTNER, Michael (2010). «Interventionism and Epiphenomenalism». *Canadian Journal of Philosophy* 40: pp. 510–529.
- BAUMGARTNER, Michael (2013). «Rendering Interventionism and Non-Reductive Physicalism Compatible». *Dialectica* 67: pp. 1–27. DOI: 10.1111/1746-8361.12008.
- BAUMGARTNER, Michael (2017). «The Inherent Empirical Underdetermination of Mental Causation». *Australasian Journal of Philosophy* 96: pp. 335–150. DOI: 10.1080/00048402.2017.1328451.
- BONTLY, Thomas D. (2002). «The Supervenience Argument Generalizes». *Philosophical Studies* 109: pp. 75–96.
- CRAVER, Carl F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Oxford University Press.
- ERONEN, Markus (2012). «Pluralistic Physicalism and the Causal Exclusion Argument». *European Journal for the Philosophy of Science* 2: pp. 219–232.
- ERONEN, M. & BROOKS, D. S. (2014). «Interventionism and Supervenience: A New Problem and Provisional Solution». *International Studies in the Philosophy of Science* 28: pp. 185–202.
- GEBHARTER, Alexander (2017). *Causal Nets, Interventionism, and Mechanisms: Philosophical Foundations and Applications*. Berlin: Springer.
- HAUSMAN, Daniel & WOODWARD, James (1999). «Manipulation and the Causal Markov Condition». *Philosophy of Science* 50: pp. 521–583. DOI: 10.1086/425235
- HITCHCOCK, Cristopher (2018). «Causal Models». En: *Stanford Encyclopedia of Philosophy*, editado por Edward N. Zalta. Stanford, CA: Stanford University. Publicado por la primera vez el 7 de agosto de 2018. Consultado el 22 de diciembre de 2019. Disponible en: <https://plato.stanford.edu/entries/causal-models/>.
- HODGKIN, Alan L. & Huxley Andrew F. (1952). «A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve». *Journal of Physiology* 117: pp. 500–544.
- HOLLAND, Paul (1986). «Statistics and Causal Inference». *Journal of the American Statistical Association* 81: pp. 945–960.
- KIM, Jaegwon (1998). *Mind in a Physical World*. Cambridge: MIT Press.

- KIM, Jaegwon (2005). *Physicalism or Something Near Enough*. Princeton: Princeton University Press.
- LEWIS, David (1986). «Causation (Reprint with postscripts)». *Philosophical Papers*, vol. 2: pp. 159–213. Oxford: Oxford University Press.
- LIST, Christian & Menzies, Peter (2009). «Nonreductive Physicalism and the Limits of the Exclusion Principle». *Journal of Philosophy* 106: pp. 475–502. DOI: 10.5840/jphil2009106936.
- MARRAS, Ausonio (1999). «Critical notice of Kim’s “Mind in a Physical World”». *Canadian Journal of Philosophy* 30: pp. 137–60.
- PEARL, Judea (2009). *Causality: Models, Reasoning and Inference* (2nd edition). Cambridge: Cambridge University Press.
- PEARL, Judea (2018). *The Book of Why*. New York: Basic Books.
- RAATIKAINEN, Panu (2010). «Causation, Exclusion, and the Special Sciences». *Erkenntnis* 73: pp. 349–363. DOI: 10.1007/s10670-010-9236-0.
- REUTLINGER, Alexander (2013). *A Theory of Causation in the Social and Biological Sciences*. London: Palgrave MacMillan.
- SARKAR, Sahotra (2005). *Molecular Models of Life*. Cambridge: MIT Press.
- SCHEINES, Richard (2005). «The Similarity of Causal Inference in Experimental and Non-Experimental Studies». *Philosophy of Science* 72: pp. 927–40. DOI: 10.1086/508950.
- SHAPIRO, Larry & Sober, Elliott (2007) «Epiphenomenalism. The Dos and the Don’ts». En: *Thinking about Causes: From Greek Philosophy to Modern Physics*, pp. 235–264, editado por Peter Machamer y Gereon Wolters. Pittsburg: University of Pittsburg Press.
- SPIRITES, Peter & Scheines, Richard (2005). «Causal Inference of Ambiguous Manipulations». *Philosophy of Science* 71: pp. 833–845. DOI: 10.1086/425058
- spirites, Peter; GLYMOUR, Clark & SCHEINES, Richard (2001). *Causation, Prediction, and Search* (2nd edition). Cambridge: MIT Press.
- WEBER, Marcel (2008). «Causes without Mechanisms: Experimental Regularities, Physical Laws, and Neuroscientific Explanation». *Philosophy of Science* 75: pp. 995–1007. DOI: 10.1086/594541.
- WESLAKE, Brad (*forthcoming*) Exclusion Excluded. *International Studies in the Philosophy of Science*.
- WOODWARD, James (2003). *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.
- WOODWARD, James (2008). «Mental Causation and Neural Mechanisms». En:

Being Reduced. New Essays on Reduction, Explanation, and Causation, pp. 218–262, editado por Jakob Hohwy & Jesper Kallestrup. Oxford: Oxford University Press.

WOODWARD, James (2015). «Interventionism and Causal Exclusion». *Philosophy and Phenomenological Research* 77: pp. 193–212. DOI: 10.1111/phpr.12095.

WOODWARD, James (2016). «Causation and Manipulability». En: *Stanford Encyclopedia of Philosophy*, editado por Edward N. Zalta. Stanford, CA: Stanford University. Revisión substantiva al 14 de octubre de 2016. Consultado el 22 de diciembre de 2019. Disponible en: <https://plato.stanford.edu/entries/causation-mani/>.

WOODWARD, James (2018). «Intervening in the Exclusion Argument». En: *Making a Difference*, pp. 251–268, editado por Helen Beebe; Christopher Hitchcock & Huw Price. Oxford: Oxford University Press.

YABLO, Stephen (1992). «Mental Causation». *Philosophical Review* 101: pp. 245–280. DOI: 10.2307/2185535.



Manipulating Shadows: Towards an Evidentialist (Dis)solution of the Causal Exclusion Problem

A classic problem in philosophy of mind is the possible exclusion of mental causes with respect to the physical ones. This problem, popularised by Jaegwon Kim, has undesirable consequences for any scientific discipline committed with causal claims related to supervenient properties. Focusing on issues concerning causal claims in the special sciences, I discuss in this paper the solution to the causal exclusion problem put forward by a group of authors occasionally called «evidentialists». Evidentialists consider that causal exclusion problem can be solved by appealing to the manipulationist notion of cause, originally formulated by James Woodward. Throughout the last 10 year, the evidentialist view has received some criticisms by Michael Baumgartner. My goal is to show in which respects these criticisms are correct, but do not compromise the overall validity of the evidentialist view.

Keywords: Supervenience · Causal Models · Manipulationism · Evidentialism.

Manipulando sombras: Hacia una (di)solución evidencialista del problema de la exclusión causal

Un problema clásico en filosofía de la mente es el de la posible exclusión de las causas mentales respecto a las físicas. Tal problema, popularizado por Jaegwon Kim, tiene consecuencias indeseables para cualquier disciplina científica que se comprometa con atribuir relevancia causal a propiedades supervenientes. Enfocando mi atención sobre cuestiones relacionadas con atribuciones causales en las ciencias especiales, discuto en este artículo la solución al problema de la exclusión causal propuesta por un grupo de autores ocasionalmente denominados «evidencialistas». Los evidencialistas consideran que el problema de la exclusión causal puede ser solucionado apelando a la noción de causa manipulacionista, elaborada

originariamente por James Woodward. A lo largo de los últimos 10 años, la perspectiva evidencialista ha recibido algunas críticas por parte de Michael Baumgartner. Mi objetivo es el de mostrar en que aspectos estas críticas son acertadas, pero no comprometen substancialmente la validez de la perspectiva evidencialista.

Palabras Clave: Superveniencia · Modelos causales · Manipulacionismo · Evidencialismo.

LORENZO BARAVALLE es actualmente miembro del Departamento de Historia y Filosofía de la Ciencia de la Facultad de Ciencias de la Universidad de Lisboa (Portugal), así como del Centro de Ciencias Naturales y Humanas de la Universidad Federal del ABC (Brasil). Obtuvo su Doctorado en Filosofía [≈ PhD] en la Universidad de Barcelona y en la Universidad Rovira i Virgili (España). Está interesado en los intentos de generalización de la teoría evolutiva a dominios no estrictamente biológicos, tales como la cultura, la computación, y la epistemología. Ha publicado sobre tópicos relacionados con la estructura teórica de la teoría evolutiva y las explicaciones evolutivas.

INFORMACIÓN DE CONTACTO | CONTACT INFORMATION: Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências; Faculdade de Ciências, Universidade de Lisboa, 1749-016, Lisboa, Portugal. e-mail (✉): lbaravalle@fc.ul.pt · iD: <https://orcid.org/0000-0001-6299-6403>

HISTORIA DEL ARTÍCULO | ARTICLE HISTORY

Received: 16-May-2019; Accepted: 26-November-2019; Published Online: 27-December-2019

COMO CITAR ESTE ARTÍCULO | HOW TO CITE THIS ARTICLE

Baravalle, Lorenzo (2019). «Manipulando sombras: Hacia una (di)solución evidencialista del problema de la exclusión causal». *Disputatio. Philosophical Research Bulletin* 8, no. 11: pp. 355–380.

© Studia Humanitatis – Universidad de Salamanca 2019