

LA TEORÍA DRETSKEANA DE LA CAUSACIÓN MENTAL ANTE LA EXPLICACIÓN PSICOLÓGICA

Agustín VICENTE
Universidad de Barcelona

Durante la última década, Fred Dretske ha defendido una teoría naturalizadora de lo mental que intenta dar cuenta de uno de los hechos más sobresalientes de las propiedades mentales, a saber, que tales propiedades tienen eficacia causal. Se trata de una teoría cuyas líneas maestras son:

- (a) el compromiso con un externismo que cabría denominar de «sentido común»,
- (b) la explicación de las nociones semánticas en términos de nociones no semánticas, en concreto, en términos del concepto matemático de información, y
- (c) la afirmación de que las propiedades mentales actúan como «causas estructurantes» del comportamiento.

En este breve artículo me centraré en los problemas que rodean al último de estos puntos, y dejaré de lado otros problemas que pueda tener la teoría naturalista de Dretske. Para ello, desarrollaré en primer lugar, pero muy escuetamente, su teoría de la causación mental como causación estructurante, presentaré a continuación dos problemas ya clásicos (y plausiblemente irresolubles) para tal teoría, y finalmente propondré una manera de sortear el segundo de los problemas reseñados.

1. Cabe decir que la teoría de Dretske tiene dos polos explicativos, dos tipos de hechos que intenta cubrir: por una parte, trata de desarrollar una explicación naturalista del contenido mental recurriendo a la noción de información; por la otra, procura dar cuenta de la eficacia causal de las pro-

piedades mentales. Lo que Dretske afirma, en ese respecto, es que los eventos mentales causan *estructurantemente* lo mismo que los eventos neurológicos causan *desencadenantemente*. Cuando proporcionamos una explicación en términos mentalistas, estamos dando cuenta del por qué del comportamiento que explicamos, en contraste con el cómo que nos proporciona la explicación de la física (o, en este caso, de la neurología).

La diferencia entre un tipo de explicación y otro se hace patente recurriendo a uno de los ejemplos del propio Dretske: imaginemos que un terrorista coloca una bomba adosada a los bajos de un coche, bomba que explota al accionar la llave de contacto. Tras la explosión, preguntamos, en primer lugar, por la causa de la deflagración, y se nos explica que la bomba estalló porque la víctima entró en el coche y giró la llave, accionando de esa forma el detonador de la bomba; después, insatisfechos, inquirimos por la causa del proceso giro de llave-causa-explisión de bomba, y es entonces cuando se nos habla del terrorista y de cómo colocó una bomba en el coche. El primer tipo de causación es causación desencadenante, mientras que el segundo es la citada causación estructurante.

Siguiendo este patrón, cabe explicar la actuación causal de las propiedades mentales del siguiente modo: tomemos C, una estructura neurológica localizada en un organismo dado, y M el comportamiento (movimiento corporal) a explicar, por ejemplo, un movimiento de huida. Supongamos que C se activa ante —indica— una cierta propiedad del entorno, F, propiedad de cuya presencia le conviene estar informado al organismo, pues, pongamos, F es uno de sus predadores. Dadas estas condiciones, C puede ser *reclutado* para causar el movimiento de huida M cada vez que F está presente en el entorno. Así las cosas, el hecho de que C indique F, nos explica por qué las propiedades neurológicas de C causan —desencadenantemente— M: en algún momento de su historia, C fue reclutada para causar M en virtud de que indicaba F.

Dretske distingue entre indicación y representación. En el momento anterior al reclutamiento, la estructura C sólo indica F, i.e. se correlaciona con o es causada por F. Eso quiere decir que antes del reclutamiento no existe la posibilidad del error, no hay indicación errónea. La normatividad se adquiere con el reclutamiento: a partir de ese momento, C adquiere una nueva propiedad, la propiedad teleológica de tener la función de indicar, que no es otra que la propiedad de representar.

El primer problema serio que ha de afrontar la teoría dretskeana consiste en su incapacidad para procurar un trabajo causal a las propiedades representacionales. Si la propiedad de tener la función de indicar surge del reclutamiento, entonces no puede tener parte en él. Ese trabajo causal lo realiza, únicamente, la propiedad de indicar. Y eso quiere decir que no son las propiedades mentales, sino propiedades carentes de normatividad, las que causan estructurantemente las conductas.

El segundo problema, desvelado por gran cantidad de críticos (cf. Cummins (1991), Dennett (1991), Horgan (1991), Rudder Baker (1991)), al que la teoría de Dretske no parece ser capaz de responder, es el que podemos llamar «problema del aquí-y-ahora». Nuestra intuición, dicen los críticos, no es tan sólo que nuestras creencias y deseos causan nuestros comportamientos, sino que lo hacen «aquí-y-ahora», es decir, que nuestros comportamientos tienen a creencias y deseos como antecedentes causales *locales*. Sin embargo, la teoría de Dretske ubica la eficacia causal de las propiedades indicadoras en un momento, seguramente bastante alejado, del pasado, cuando se dio aquel proceso de reclutamiento, por ejemplo, al finalizar el proceso de aprendizaje de un concepto.

2. Como digo, centraré mi atención en este segundo problema, aunque llevado al terreno de la explicación. Por lo que yo veo, el problema metafísico que se le plantea a Dretske es irresoluble: su teoría no puede proporcionar causación mental aquí-y-ahora, ya que explícitamente afirma que tal causación es operativa en el pasado. Lo único que cabe hacer es, como se dice, morder la bala, y negar la fuerza de la intuición. Cabe decir: «yo no tengo tan claro que mis creencias y deseos causen mis conductas de ese modo aquí-y-ahora».

Sin embargo, es un hecho empírico, y no intuitivo, que utilizamos la psicología para proporcionar explicaciones causales desencadenantes de los comportamientos. Es cierto que a veces preguntamos explícitamente por causas estructurantes. En ocasiones uno quiere saber qué suceso psicológico estructura el proceso que desemboca en un cierto comportamiento; conoce la causa desencadenante, un pensamiento que ha cruzado su cabeza y le ha obligado a hablar en voz un poco más alta de lo habitual, pero desea saber por qué acostumbra a sucederle esas cosas, en lugar de llevarlo todo con más calma, como hacen otros. Pero esto no es la norma, y de hecho, el que hayamos citado un pensamiento como causa desencadenante del comportamiento da idea de que son muchas las veces que la psicología proporciona explicaciones desencadenantes. No en vano, hablamos de mecanismos cognitivos.

De acuerdo a lo que hemos venido diciendo hasta el momento, sin embargo, las propiedades que maneja la psicología no son, estrictamente, los antecedentes causales de las conductas. Se trata sólo de propiedades que causan estructurantemente, y en el pasado, tales conductas. ¿Cómo podemos reconciliar ambos hechos, de una parte explicación causal desencadenante y de otra causación estructurante operativa en el pasado?

Creo que Fodor (1994) ha procurado una solución a este problema, solución que podemos expresar de la manera siguiente:

(*) si, *ceteris paribus*, todos y sólo los individuos que instancian una propiedad P instancian alguna propiedad Q tal que (la instanciación de) Q causa (la instanciación de) M, entonces la ley «los Ps causan M» es aceptable.

Esta solución nos permite, por ejemplo, aceptar la ley «los caballos procrean cuando se aparean entre sí», a pesar de que la propiedad de ser un caballo, i.e. de haber nacido de caballos, es epifenoménica a la hora de producir nuevos caballos, y de que, en principio, (a) cabe encontrar o fabricar un espécimen que haga todo lo que hace un caballo sin que en su origen figuren otros caballos y (b) hay caballos que son incapaces de procrear.

Dado que la Naturaleza nos ofrece una correlación fiable entre caballos e individuos con los poderes causales de los caballos (i.e. todos y sólo los individuos que instancian la propiedad de ser caballo suelen instanciar tales y cuales poderes causales), podemos despreocuparnos de tales eventualidades al elaborar nuestra ciencia, y hacer uso de la propiedad *caballo* en lugar de una suerte de *narrow-caballo*, o de las micro-propiedades que causen cualesquiera efectos que adjudicamos a los caballos. Este uso de la propiedad *caballo* ha de entenderse como afirmando que hay alguna propiedad que instancian los individuos que poseen la propiedad de ser un caballo, tal que es responsable de los efectos a explicar.

Jackson y Pettit (1990), por su parte, han desarrollado un modelo de explicación —modelo de explicación por programación— que permite aceptar las explicaciones causales funcionales, aun cuando se acepta que las propiedades funcionales son inertes. El uso de predicados funcionales, afirman, resulta explicativo porque nos informa de que alguno de sus realizadores —y no otra propiedad cualquiera— causó el efecto y de que el «sendero causal» que siguió la instanciación de esa propiedad fue el marcado por el predicado funcional.

Dicho esto, la propuesta de Fodor puede entenderse como una ampliación del modelo de Jackson y Pettit a aquellas propiedades que se correlacionan contingente aunque fiablemente con las propiedades responsables de los efectos a explicar. Si asumimos que las propiedades tienen sus poderes causales esencialmente, entonces, en principio, el modelo de explicación por programación quedaría restringido a las propiedades que se correlacionan por necesidad con las propiedades causalmente eficaces —i.e. las propiedades funcionales—; las ideas de Fodor nos permiten ampliar el modelo a propiedades como la de *caballo*, que se correlaciona contingentemente con las propiedades físicas que causan cosas como la generación de nuevos caballos.

3. Así pues, parece que podemos hacer uso de propiedades representacionales para explicar desencadenantemente conductas siempre y cuando se correlacionen fiablemente con las propiedades neurológicas responsables de tales conductas. ¿Cómo sabemos que estamos ante un caso de correlación fiable? En principio cabe imaginar que una determinada conducta fuera causada por propiedades que no son sensibles a la información, o que son sensibles a otro tipo de información. En ocasiones, de hecho, ocurre así, y se mueve un brazo «por ninguna razón» o «por equivocación». Pero, claramente, como diría Fodor, se trata de hechos accidentales que cubren las cláusulas *cp*. Los mundos en los que estas cosas ocurren a modo de norma por las razones que sean, nos caen demasiado lejos. Por otra parte, nuestro mundo tampoco recoge muchos casos de propiedades representacionales instanciadas en sujetos por otro lado carentes de propiedades que causen conductas.

La razón de que exista esta correlación fiable radica en que, presumiblemente, si no fuera porque instanciaciones pasadas de las propiedades representacionales causaron la conexión entre propiedades neurológicas intrínsecas y movimientos corporales, estas propiedades intrínsecas no causarían esos movimientos corporales.

Para concluir: utilizamos propiedades representacionales para explicar efectos de los que no son responsables porque, *ceteris paribus*, siempre que se da una propiedad responsable de esos efectos, se da una propiedad representacional, y las propiedades representacionales nos resultan epistémicamente más accesibles, son más adecuadas y proporcionadas a los efectos a explicar, y ponen a nuestra disposición generalizaciones que de otro modo perderíamos. Todo ello, a pesar de que su eficacia causal tiene lugar en el pasado y son epifenoménicas en el aquí-y-ahora.

Referencias

- BAKER, L. R. (1991): Dretske on the Explanatory Role of Belief. *Philosophical Studies*, 63, pp. 99-111.
- CUMMINS, R. (1991): The Role of Mental Meaning in Psychological Explanation. In McLaughlin (ed.) 1991.
- DENNETT, D. (1991): Ways of Establishing Harmony. In McLaughlin (ed.) 1991.
- DRETSKE, F. (1988): *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA.: MIT Press.
- DRETSKE, F. (1991): Replies. En McLaughlin (ed.) 1991.
- DRETSKE, F. (1993): Mental Events as Structuring Causes of Behavior. En Heil y Mele (eds.) *Mental Causation*. Oxford: Clarendon Press.
- DRETSKE, F. (1994): Reply to Slaten and García-Carpintero. *Mind and Language*, 9, pp. 203-208.
- DRETSKE, F. (1996): How Reasons Explain Behavior: Reply to Melnyk and Noordhof. *Mind and Language*, 11, pp. 223-229.
- FODOR, J. (1994): *The Elm and the Expert*. Cambridge, MA.: MIT Press.
- HORGAN, T. (1991): Actions, Reasons, and the Explanatory Role of Content. En McLaughlin (ed.) 1991.
- JACKSON, F. y PETTIT, P. (1990): «Program Explanation: A General Perspective», *Analysis* 50: 107-117.
- McLAGHLIN, B. (ed.) (1991): *Dretske and his Critics*. Cambridge, MA.: MIT Press.