

## ARGUMENTO CAUSAL, EXPLICACIÓN PSICOLÓGICA Y MODELOS NO-CAUSALES: ¿PUEDEN LAS PROPIEDADES EXTERNISTAS SER EXPLICATIVAMENTE RELEVANTES PARA LA CONDUCTA?

Causal argument, psychological explanation, and non-causal models: Can externalist properties be explanatorily relevant to behavior?

Mario Felipe Salinas Mejias<sup>1</sup>

Universidad Alberto Hurtado, Santiago de Chile

msalinasmejias@gmail.com

### Resumen

En este artículo examino el argumento causal defendido por algunos filósofos que sostienen que sólo las propiedades intrínsecas o internas del sujeto cognitivo son relevantes para explicar la conducta, basándose en la relación entre propiedades causalmente eficaces y explicativamente relevantes. Aunque usualmente se acepta que las propiedades causalmente eficaces son intrínsecas, sostendré que el argumento causal no logra cumplir su propósito. Esto se debe a que las propiedades extrínsecas o externas pueden ser igualmente explicativas, ya que las propiedades que determinan la relevancia explicativa pueden caracterizarse, no solo en términos causales, sino también en términos no causales. Hacia el final del trabajo, y a partir de nuevos modelos explicativos provenientes de las neurociencias y matemáticas, mostraré cómo los modelos no causales podrían ser una alternativa interesante para explicar la conducta en ciertos casos.

*Palabras clave:* argumento causal, explicación de la conducta, externismo, internismo, explicación no causal.

### Abstract

---

<sup>1</sup> Licenciado y Magíster en Filosofía por la Universidad Alberto Hurtado. Doctorando en Filosofía por la misma casa de estudios.  
<https://orcid.org/0009-0009-8064-0182>.

In this article, I examine the causal argument defended by some philosophers, who argue that only the intrinsic or internal properties of the cognitive subject are relevant for explaining behavior, based on the relationship between causally efficacious and explanatorily relevant properties. Although it is commonly accepted that causally efficacious properties are intrinsic, I will argue that the causal argument fails to fulfill its purpose. This is because extrinsic or external properties can be equally explanatory, as the properties determining explanatory relevance can be characterized not only in causal terms but also in non-causal terms. Towards the end of the paper, drawing from new explanatory models emerging from neuroscience and mathematics, I will show how non-causal models could be an interesting alternative for explaining behavior in certain cases.

*Keywords:* causal argument, explanation of behavior, externalism, internalism, non-causal explanation.

*Fecha de Recepción:* 10/03/2025 — *Fecha de Aceptación:* 07/05/2025

## 1. Introducción

Las reflexiones teóricas acerca de significado y a elementos semánticos relacionados tales como nombres propios, nombres de clase natural, indexicales, descripciones definidas y referencia, dieron lugar rápidamente a consideraciones que extendieron los límites de la filosofía del lenguaje, en su uso restringido —es decir, solo consideraciones lingüísticas—, hacia filosofía de la mente, filosofía de la psicología, filosofía de las ciencias y ciencias cognitivas. Las discusiones entre internismo-externismo acerca del significado, a través del análisis de los conceptos mencionados, se trasladaron hacia el análisis de los estados mentales (EM) en términos de contenido, particularmente en lo referente a sus condiciones de individuación. Esto fue influenciado, principalmente, por el desarrollo del externismo semántico en un sentido más amplio—es decir, a elementos más allá de lo puramente

lingüístico—, específicamente por los trabajos de Kripke (1980) y su análisis sobre los nombres propios; Kaplan (1989) y su análisis sobre indexicales; Putnam (1975) sobre el análisis de nombres de género natural —externalismo físico o natural— y, especialmente, Burge (2007) y su análisis sobre los factores sociolingüísticos implicados en el significado (externismo social). Precisamente a través del trabajo de Putnam y Burge se sentaron las bases de lo que se puede denominar externismo de contenido (EC), lo cual tuvo inmediatas consecuencias conceptuales respecto del análisis del contenido de los EM.

En ese sentido, los planteos acerca de la individuación del contenido de los EM han generado un extenso debate en filosofía de la mente, en el contexto de las discusiones entre internismo y externismo de contenido. Por un lado, el internismo de contenido (IC) puede describirse como la tesis según la cual el contenido de los EM de S depende totalmente de propiedades intrínsecas o internas de S, tales como alguna propiedad cerebral o alguna perteneciente a los sistemas internos (*e.g.*, a otras partes del sistema nervioso). Por otro lado, el externismo de contenido (EC) puede describirse como la tesis según la cual el contenido de los EM depende parcialmente de factores externos, propiedades relacionales o históricas del individuo, tales como la constitución física de las clases naturales o el rol de las instituciones sociales que inciden en la determinación del contenido, de modo que el contenido es fijado o determinado por esas propiedades.

En este contexto, un grupo de filósofos<sup>2</sup> han formulado una objeción contra el EC. Su objeción consiste en afirmar que: si el externismo es el caso, entonces los contenidos de los EM de S no pueden figurar como propiedades explicativas de la conducta de S. Uno de los aspectos que está a la base de esa objeción es el hecho de que existe un compromiso acerca de la explicación y la causalidad. Usualmente, los defensores de esta tesis se comprometen con lo que se puede denominar *intuición internista*, una posición acerca de la causalidad y la

---

<sup>2</sup> Esta asunción ha sido ampliamente defendida en la literatura filosófica cómo es posible encontrar en Newman (2002), Stich (1983), McGinn (1989), Segal (2000), Yablo (1997), Kim (1982, 1998).

explicación, según la cual solo las propiedades intrínsecas o internas son epistémicamente relevantes para explicar la conducta de S.

Esta idea depende de dos consideraciones. Por un lado, depende de una concepción relativamente amplia en filosofía de las ciencias acerca de la explicación, ya que en ese sentido se asume que existe un vínculo muy estrecho entre explicación y causalidad, por lo que la explicación tiene una estructura causal, de modo que explicar un fenómeno consiste en dar cuenta de la información causal de su ocurrencia. Y, por otro lado, la objeción internista depende de un argumento que puede denominarse causal, el cual contiene en su base un principio bastante extendido en filosofía de la mente: el *principio de localidad*<sup>3</sup>. Este principio establece que las propiedades causalmente eficaces son intrínsecas, internas o locales, debido a que estas son las propiedades que están involucradas en los mecanismos causales. Teniendo estos dos elementos conceptuales a la base, los defensores de la posición internista afirman que el contenido externamente individuado no constituye una propiedad causalmente eficaz, por lo cual no puede ser explicativamente relevante. Por lo tanto, solo las propiedades internistas pueden ser explicativamente relevantes, dado que solo estas propiedades tienen eficacia causal.

En este trabajo argumentaré que esta concepción es errónea y que las propiedades externalistas son genuinamente explicativas, ya que el vínculo entre explicación y causalidad que asume el internalista depende de una noción particular de causalidad que no es evidente. En cambio, sostendré que las propiedades metafísicas que permiten identificar los factores explicativamente relevantes no solo se aplican a causas o relaciones causales, también pueden aplicarse a contextos no causales de la explicación, debido a que hay un tipo de relación

---

<sup>3</sup> Este principio aparece en la literatura alternativamente con el nombre de causalidad local. La idea central que transmite este principio es que no hay causalidad a distancia y ha sido enfatizado de diferentes formas en la literatura, en el contexto de esta discusión. Por ejemplo, Kim (1998, p. 205) afirma que la causa próxima del comportamiento físico debe ser local: “una cierta serie de eventos neurales que se originan en mi sistema nervioso central que causa la contracción de músculos apropiados y tal”. Kallestrup (2012) afirma “las causas del comportamiento deben residir dentro de los cuerpos de los agentes. Lo que hace que el brazo de María se mueva hacia un vaso de agua son propiedades neurofisiológicas intrínsecas”.

metafísica que subyace al poder explicativo, que son las relaciones contrafácticas. En suma, el internalista nos debe un argumento adicional para sostener su punto. Para ello primero introduciré los elementos conceptuales respecto de la relación entre explicación y causalidad; en segundo lugar, me referiré a la relación entre eficacia causal y relevancia explicativa; en tercer lugar, expondré el argumento causal que utilizan los internistas; en cuarto lugar presentaré argumentos para resistir la objeción internalista; y por último, esbozaré una manera en que el externalista podría explicar la conducta sin apelar al modelo causal tradicional.

## **2. Explicación Causal**

El vínculo entre causalidad y explicación tiene profundas raíces en la literatura filosófica, no obstante, describir qué es una explicación causal es una cuestión controvertida que ha dado lugar a una gran variedad de enfoques. Sin embargo una de las propuestas de explicación causal más influyentes en el contexto de esta discusión, y en filosofía de las ciencias, fue la de Salmon (1984, 1989), quien desarrolló un modelo causal mecánico (MCM). Salmon desarrolla una teoría causal de la explicación, en términos probabilísticos, en donde explicar un fenómeno es mostrar cómo los acontecimientos y regularidades estadísticas encajan en la red causal del mundo (Galavotti, 2005). Un aspecto central de la teoría es la noción de proceso causal, en donde dicho proceso es representado por interacciones físicas espacio-temporales susceptibles de transmitir estructura, información e influjo causal. Uno de los elementos que constituyen el núcleo de los procesos causales es lo que Salmon denomina “marca”, que a su vez lo distingue de los pseudoprocesos. Los procesos causales tienen la capacidad de transferir o transmitir una señal o marca a otros procesos. Estos son caracterizados por transmitir influencia causal y la información o marca que es transmitida, corresponde a magnitudes físicas, pudiendo alterar las estructuras de los procesos que participan en las interacciones causales. Los pseudoprocesos son incapaces de transmitir señales o marcas.

Uno de los problemas conceptuales al cual se enfrenta este tipo de teoría radica precisamente en el criterio de distinción entre procesos y pseudoprocesos. Dado que el concepto de “marca” tiene una carga contrafáctica —qué sucedería si a la relación de  $x \rightarrow y$  se le introduce  $M(\text{marca})$ —, la principal dificultad es que encontremos eventos azarosos donde haya causalidad y no explicación. En otras palabras, el criterio para distinguir entre procesos y pseudoprocesos puede no recoger las propiedades explicativamente relevantes de los procesos causales<sup>4</sup>.

Tras esto, Salmon decidió modificar su teoría y, en vez de la noción de marca para la determinación o identificación de un proceso genuinamente causal, se concentró en comprender los procesos causales en términos de cantidades conservadas, tales como masa-energía, momento lineal y carga. De este modo, un proceso causal es aquel que participa en una interacción causal y es capaz de modificar las estructuras. Es una línea de mundo (*world-line*) de un objeto que transmite una cantidad mayor a 0 en cada punto del espacio-tiempo de la trayectoria de ese objeto. Entonces, una interacción causal solo se da cuando dos líneas de mundo colisionan modificando el valor de las cantidades conservadas entrantes y salientes.

A pesar de que existen varias dificultades dentro de este marco, y abordarlas está más allá del alcance de este artículo, la relación entre causalidad y explicación ha sido muy influyente en filosofía de las ciencias, y en algunas discusiones contemporáneas en filosofía de la mente. En particular, dentro del contexto del debate entre internismo/externismo de contenido, ha persistido una imagen de la explicación que tiene un fuerte vínculo con el tipo

---

<sup>4</sup> Hitchcock (1995, p. 310) presiona sobre este punto con un ejemplo ilustrativo. Considere las bolas de billar: suponiendo que la bola blanca, la bola 8 y la butaca (*corner pocket*) se encuentran perfectamente alineadas, un tiro preciso de un jugador experto resultará en la bola 8 en la butaca. No obstante, antes del tiro el jugador marcó su taco con un poco de tiza azul, por lo que parte de esa tiza se desprendió de la bola blanca al momento de ser impactada. Hitchcock afirma que la explicación de la caída de la bola 8 en la butaca puede ser descrita invocando a las condiciones iniciales como: el momento lineal impartido por la bola blanca antes del impacto, y la mancha de tiza azul. No obstante, aquí surge un problema, porque tanto el momento lineal de la bola blanca, como la mancha de tiza azul, son marcas transmitidas por el proceso causal, sin embargo la mancha de la tiza azul es irrelevante en la explicación de la caída de la bola 8. Para una discusión más a fondo sobre este asunto ver Woodward y Ross (2022).

de modelo salmoneano, un tipo de explicación en donde la causalidad es fundamental, y además se entiende en un sentido *óntico*.

Ahora bien ¿cómo se articula esto en el debate entre IC y EC? El vínculo entre causalidad y explicación suele instalarse en el discurso filosófico del siguiente modo: por un lado, las explicaciones de la conducta apelan generalmente al marco de la psicología *folk* y este marco explica la conducta en referencia a los contenidos o atribuciones intencionales de un sujeto cognitivo en las que dicho sujeto tiene una cierta actitud proposicional que, típicamente, refiere a la conjunción entre deseos y creencias. De este modo, la explicación, por ejemplo, de la conducta de María de ir por un café al restaurante de la esquina, se explica por el deseo de María de beber café y de su creencia de que en el restaurante obtendrá café. La explicación de la psicología *folk* surge a partir de la asignación de estados intencionales, en este caso, la creencia y el deseo de María. Por otro lado, una de las intuiciones acerca de la causalidad referida a nuestros propios contenidos mentales apela a que existe una relación estrecha entre dichos contenidos y nuestra conducta, de modo que nos comportamos de acuerdo con lo especificado por nuestros contenidos mentales. Y el problema de la causación mental puede especificarse mediante la pregunta de cómo los contenidos se relacionan con la conducta. Puesto en otros términos, si la conducta se va a explicar en términos causales y el marco explicativo es la psicología *folk* que apela precisamente a creencias y deseos, específicamente al contenido externamente individuado de esos estados, entonces ¿cómo es que ese contenido puede hacer una diferencia causal?

Los defensores de la intuición internista consideran que en las explicaciones causales la causalidad es local, es decir, lo que explica lo hace en virtud de una propiedad de la estructura interna del agente o una propiedad intrínseca de este. Por ejemplo, lo que explica la conducta de María yendo al restaurante es el mecanismo físico interno que permite la acción, el movimiento. Lo que explica la conducta de María de subir las escaleras también son hechos intrínsecos acerca de ella.

Entonces podemos visualizar dos premisas relacionadas con la explicación que conducen a una conclusión anti-externista: (I) la explicación de la conducta se establece a

partir de las creencias y deseos; (II) la concepción causal de la explicación implica que los contenidos de los EM explican la conducta solo si la causan.

El problema de la explicación de la psicología *folk* con el argumento causal surge cuando se considera la plausibilidad de la tesis externista: la idea de que el contenido de los estados mentales depende constitutivamente de las propiedades relacionales, históricas y externas. De este modo, las creencias y los deseos dependerían no solo de los hechos intrínsecos acerca de ese individuo, como las propiedades neurobiológicas o referidas a los sistemas internos, sino también de hechos externos como la constitución física de las clases de género natural, como “agua”, o el rol de las instituciones sociales en la determinación del contenido o del pensamiento.

Los principales argumentos a favor del externismo de contenido provienen de los trabajos de Putnam (1975) y Burge (2007), quienes argumentan cómo esos hechos externos, de naturaleza física en caso de Putnam y social en el caso de Burge, tienen un rol en la constitución del contenido. Una de las ventajas de la tesis externista es que puede dar cuenta de cómo las propiedades naturales/sociales tienen un rol importante respecto de los contenidos de los estados intencionales, como deseos y creencias. En el ámbito de la conducta, esto es sumamente relevante, porque a partir de la tesis externista se podría dar cuenta eventualmente de cómo el entorno influye en la conducta y de cómo la causación mental no requiere la intervención de propiedades no naturales.

Teniendo eso en consideración, la objeción internista consiste en afirmar que, si el externismo es el caso, es decir, si el contenido de los EM depende constitutivamente de propiedades externas, entonces el externismo es incompatible con las explicaciones causales en el ámbito de la conducta. Esto es porque, en su caracterización externista, los estados intencionales como creencias y deseos son “causalmente ineficaces” y, por ende, no son relevantes para la explicación de la conducta. A continuación profundizaré en esto.

En una serie de artículos Dretske (1988, 2015) ilustra el problema de la causación mental con una famosa analogía. Nos pide que imaginemos una máquina expendedora, esta máquina funciona introduciendo dinero para obtener ciertos objetos como dulces, cigarrillos,

etc. Para que el intercambio ocurra en la máquina debe introducirse el tipo correcto de dinero, es decir, el tipo de objeto que hace que la máquina se comporte de un cierto modo, intercambiando ese dinero por el tipo de objetos descrito. Ahora bien, para que ciertos objetos cuenten como dinero, es decir tengan la propiedad de ser dinero, ya sean monedas o billetes, debe haber una relación extrínseca que esté operando, dado que dicha propiedad depende de hechos externos a las monedas o billetes. Por ejemplo, el valor monetario puede estar determinado por las políticas tributarias de los gobiernos, entre otros factores. Esto quiere decir que las propiedades intrínsecas de las monedas o los billetes no determinan el valor del dinero. Podemos pensar ahora la relación entre lo mental y lo físico, en términos de la relación entre el dinero y las máquinas expendedoras.

En este relato, Dretske nos invita a pensar que la mente es al comportamiento o acción, como las monedas son a las máquinas expendedoras. Es decir, si introduces las monedas correctas obtienes un determinado comportamiento de la máquina, al igual que ciertas propiedades mentales o entidades producen un comportamiento determinado en el cuerpo. Y así mismo, como el dinero obtiene su valor por estar en una relación apropiada con hechos externos, también las entidades mentales, como las creencias, obtienen su contenido debido a relaciones con hechos externos:

[...] lo que hace que el dinero sea dinero no es su carácter intrínseco, por ejemplo, la forma, el tamaño y la densidad de las monedas, sino ciertos hechos extrínsecos o relacionales sobre estas monedas (el hecho de que poseen un valor monetario), también lo que hace que una creencia sea una creencia, no es su carácter neurobiológico intrínseco, sino más bien, ciertos hechos extrínsecos al respecto: el hecho de que tiene un cierto significado o contenido, el hecho de que tiene ciertas propiedades intencionales (Dretske, 1998, p. 157)<sup>5</sup>.

---

<sup>5</sup> Todas las citas directas provenientes del idioma inglés han sido traducidas por el autor.

Entonces al tener ciertos ítems mentales, como las creencias, se produce una conducta acorde a ese ítem. Sin embargo, aquí surge una primera dificultad, si nos situamos en el contexto en el que debemos explicar qué es lo que da cuenta del comportamiento de la máquina, por ejemplo, cuando pones un billete y te da un refresco, un café, o cosas similares, parece ser que no son las propiedades relacionales, no es en este caso la propiedad de que el billete o moneda tenga un valor monetario lo que explica el comportamiento de la máquina, sino más bien, son propiedades intrínsecas las que lo hacen, propiedades tales como la densidad y la forma de los billetes, los patrones de la tinta y cosas similares son las que explican el comportamiento de la máquina, ya que son el tipo de cosas que las máquinas pueden detectar. Por el contrario, estas máquinas no detectan las propiedades como la historia de estos objetos o la propiedad de ser dinero.

Si consideramos lo anterior, entonces eso nos pone ante una dificultad respecto de la analogía, ya que en el relato de Dretske operamos como las máquinas expendedoras, los ítems mentales funcionan como monedas respecto del comportamiento, pero estos ítems son irrelevantes si consideramos que el contenido depende de hechos extrínsecos, nuestro organismo solo es sensible a entradas-estímulos y salidas respectivas, es decir, acciones que responden en términos intrínsecos. Considerando ese hecho, al hacer la pregunta por la conexión entre el contenido y el comportamiento, parece ser que el contenido, es decir, las creencias y los deseos, no encajan en la explicación del comportamiento porque no pueden figurar en los mecanismos causales-mentales que dan lugar a la conducta (Robb & Heil, 2021). Si esto es así, entonces es difícil ver cómo las propiedades externamente individuadas pueden figurar en las explicaciones psicológicas.

Esta asunción ha sido defendida por algunos filósofos y filosofas de manera directa o indirecta. Algunos ejemplos en esta dirección:

- (i) “Tesis explicativa: los estados psicológicos internos son los únicos estados que la teoría psicológica necesita invocar para explicar el comportamiento humano, los únicos estados necesarios para la psicología” (Kim, 1982, p. 59).

- (ii) Las propiedades relacionales son semánticas o extrínsecas mientras que las propiedades causales involucradas en la producción del comportamiento son propiedades intrínsecas no relacionales del organismo. Si los estados internos están implicados en la causalidad del comportamiento, todo el trabajo causal se hace por sus propiedades sintácticas (Kim, 1991, p. 55).

McGinn (1989) comienza afirmando la tesis externalista según la cual los contenidos de los EM, externamente individuados, se determinan o identifican por referencia a entidades que están fuera del cráneo del sujeto. Y existe una relación de correspondencia —indica— entre los estados intrínsecos y aquellas entidades que se ubican fuera del sujeto. Y luego señala esto respecto de las explicaciones psicológicas:

- (iii) [...] Para ser causales, entonces estas relaciones de correspondencia no pueden ser implicadas en la transacción causal que se informa. Esto se debe a que lo que sucede en el nexo causal es local, próximo e intrínseco: las características de la causa que conducen al efecto deben estar ahí donde tiene lugar la interacción causal. [...] La causalidad es lo mismo con los cerebros que con las bolas de billar. Sus efectos dependen de las características locales de esas entidades (McGinn, 1989, p. 133).

A pesar de motivar el externismo, McGinn se compromete con la idea de que las propiedades causalmente eficaces, o las que hacen diferencias causales, son aquellas que directamente deben figurar en las explicaciones psicológicas. Ya sea directa o indirectamente, este grupo de filósofos se comprometen con la idea de que (i) los contenidos externamente individuados no son causalmente eficaces y (ii) solo las propiedades intrínsecas pueden figurar genuinamente en las explicaciones psicológicas.

### **3. Relevancia Causal**

La objeción internista en torno a la capacidad explicativa del externismo radica por un lado, en la noción de relación causal, esto es, qué tipo de entidades se relacionan causalmente en un evento determinado y, por otro lado, en el funcionamiento de la explicación en el ámbito de la conducta. Con respecto al primer punto, los internistas van a ser muy claros; las entidades se relacionan causalmente en virtud de sus propiedades intrínsecas. Supongamos que dos eventos, A y B, se relacionan causalmente. Para que exista de hecho una relación causal tanto A como B deben instanciar propiedades particulares respectivamente. Supongamos que el evento A consiste en una bola de billar moviéndose y el evento B en el movimiento de una segunda bola producto de la interacción con la bola A. Estos eventos se relacionan causalmente en virtud de poseer ciertas propiedades intrínsecas como, por ejemplo, la masa de A y de B. Esto responde a un punto metafísico ampliamente aceptado respecto de la causalidad: *principio de localidad*. Este principio afirma que las entidades se relacionan causalmente en virtud de estructuras y mecanismos internos.

En torno a la discusión de la conducta, esas propiedades internas o intrínsecas del individuo, en las que el poder causal de los estados mentales superviene, pueden ser mecanismos neuronales. Por ejemplo, la causa de que María esté subiendo las escaleras o yendo al restaurante reside, en último término, en propiedades internas o intrínsecas como los circuitos neuronales subyacentes a nuestras capacidades motrices.

Respecto del segundo punto, es decir, del funcionamiento de las explicaciones en el ámbito psicológico, si tenemos un enfoque causal de la explicación y un enfoque local de la causalidad, entonces los ítems explicativos deben residir en el organismo en cuestión. Es decir, dado que las propiedades intrínsecas son las que hacen el trabajo causal, también serán estas propiedades las que resulten ser explicativamente relevantes.

Entonces, resumiendo, tenemos dos intuiciones. Una intuición que se puede denominar “internista”, según la cual los únicos elementos explicativos para dar cuenta de la conducta, en términos causales, son propiedades internas o propiedades intrínsecas del individuo. Y, por otro lado, está lo que se puede denominar la “intuición externista”, según la cual la conducta se explica mejor apelando a propiedades relacionales, históricas y

ambientales del individuo. Antes de comprender cómo se relacionan las propiedades causalmente eficaces con las propiedades explicativamente relevantes, necesitamos, al menos, una idea intuitiva de qué es una propiedad causalmente eficaz.

Lo primero que se puede señalar es que cuando se establecen eventos causales, o relaciones causales, intervienen propiedades que van a ser más o menos relevantes para la producción del efecto. Una relación causal puede ser una relación entre eventos, por ejemplo, el evento A causó el evento B, o relaciones entre hechos, acontecimientos, o bien estados o propiedades. Por ejemplo, el estado A causó el estado B, o bien la propiedad A causó la propiedad B. Pero, tal como señala Kallestrup (2012), independiente de cómo se conciba esta relación, para efectos explicativos, los eventos causales están constituidos por propiedades que serán, en mayor o menor medida, relevantes para la ocurrencia del efecto, tanto si son estados, eventos, acontecimiento, hechos, etc.

Por ejemplo, imaginemos a María cogiendo un revólver y apuntando en dirección a una botella de vidrio que está en línea recta desde su posición, luego supongamos que jala el gatillo del arma y al hacerlo, se inicia el mecanismo del revólver, inmediatamente sale aire extremadamente caliente del arma, debido al efecto de la pólvora, junto con la expulsión de la bala que, al entrar en contacto con la botella, la destruye. En esa descripción del evento hay una serie de propiedades que entran dentro de la relación causal y estarán determinadas, según su relevancia, para la ocurrencia del efecto. De este modo, hay ciertas propiedades que serán relevantes o eficaces para la ejecución del evento “la destrucción de la botella”, que eventualmente son propiedades físicas como, por ejemplo, la entidad física, es decir, “la bala”, su composición, masa, velocidad, la energía liberada al impactar la estructura física de la botella, etc. Existen también en ese evento propiedades que no tendrán una relevancia causal, como por ejemplo, el lugar donde fueron construidas las balas o el arma estas propiedades no explican el hecho, evento, de la botella rompiéndose, porque las relaciones causales presumiblemente son entre entidades físicas. De este modo, esas propiedades relacionales no son relevantes en términos causales.

### **3.1 Argumento Causal**

La estructura del argumento causal puede ser descrita del siguiente modo:

P1) Normalmente las explicaciones de la conducta de S, en el ámbito de la psicología, son causales.

P2) Los contenidos de los EM intencionales de S son individuados de forma amplia (tesis externalista).

P3) Solo las propiedades intrínsecas de S, o propiedades que supervienen localmente<sup>6</sup> en S, pueden ser causalmente eficaces de la conducta de S.

C1) Los contenidos de S no apelan a propiedades intrínsecas o supervienen localmente (por P2).

C2) Los contenidos no son causalmente eficaces respecto de la conducta de S (por C1 y P3).

C3) Los contenidos de S no pueden explicar la conducta de S (por P1 y C2).

En primer lugar, P1 indica que las explicaciones de la conducta de un agente cognitivo se establecen a partir de relaciones causales. La explicación de la conducta desde la *folk psychology* se articula a partir de la relación que existe entre el contenido mental y la acción del agente cognitivo. El contenido mental, en este contexto, típicamente está compuesto por dos elementos mentales: creencias y deseos. De este modo, la explicación de la conducta de María, por ejemplo, de comprar un libro de Harry Potter en la tienda, se explica por la conjunción del deseo de leer ese libro y por la creencia de que en la tienda lo conseguirá. La relación causal entre el contenido mental, el ítem deseo más el ítem creencia y la conducta de María de ir por ese libro permite que el contenido explique dicha conducta<sup>7</sup>.

---

<sup>6</sup> El concepto de “Superveniencia” puede distinguirse en términos de “global” o “local” según el rango de comparación al que se extienden las propiedades supervenientes y las que supervienen. Por ejemplo, en el nivel de comparación individual de María y María-Gemela, el rango se articula entre propiedades físicas y mentales. Pero también puede ser global si se amplía el rango de acción de las propiedades más allá del individuo, como por ejemplo a la comparación de mundos. En este caso, la superveniencia global puede incluir al entorno dentro del conjunto de propiedades en cuestión. Para mayor profundización en el concepto de Superveniencia véase Kim (1993a, 1993b, 1993c, 1993d).

<sup>7</sup> Hay, sin embargo, un conjunto más o menos heterogéneo de filósofos que no están dispuestos a aceptar esta premisa, por ejemplo, hay quienes piensan que el llamamiento a la psicología *folk* no es lo adecuado para

En segundo lugar, P2 es la tesis de EC, que indica que la mayoría de los contenidos de los estados mentales intencionales son individuados en sentido amplio, ya que los factores que determinan la constitución de estos son externos, relacionales o alternativamente se puede decir que “supervienen globalmente”. En tercer lugar, P3 indica que sólo los valores, rasgos o propiedades intrínsecas o internas — aquellas que supervienen localmente— son relevantes en términos causales, es decir, sólo esas propiedades poseen eficacia causal. En cuarto lugar, C1 se deriva de P2, ya que es una explicitación del compromiso externalista de que los contenidos mentales están constituidos parcialmente por rasgos relacionales o ambientales y, por lo tanto, no involucran propiedades intrínsecas, o internas, de forma constitutiva. C2, la negación del poder causal de los estados mentales está implicada lógicamente por C1 (que los estados mentales no están constituidos por propiedades intrínsecas) y P3 (qué sólo las propiedades intrínsecas son causalmente eficaces). Y C3 está lógicamente implicada por P1 (que las explicaciones son causales) y C2.

¿En qué radica la fuerza de este argumento? Parece ser que el núcleo del argumento depende de P1, P3 y C2. P1 hace un punto acerca de la naturaleza de la explicación, impone una restricción de orden epistémico al indicar que la estructura de la explicación se determina en términos causales. P3 enfatiza sobre un punto metafísico, la idea de que las propiedades causalmente eficaces son intrínsecas y C2 es una consecuencia directa de ella, el contenido entendido en términos externalista no posee eficacia causal. Hay, por lo tanto, una transición de lo metafísico a lo epistémico, de qué propiedades poseen eficacia causal, a qué propiedades son explicativas o no lo son, dado que la conclusión C3 es una restricción epistémica. Por lo tanto, conviene analizar en profundidad la base conceptual de cada una de estas premisas.

---

explicar el comportamiento, a veces se expresa como un rechazo al contenido mental como base conceptual para estudiar la conducta. Ejemplos de esta perspectiva pueden considerarse los proyectos eliminativistas y los enactivistas en diferentes versiones.

Un aspecto central que considerar sobre P3, es el hecho de que está apoyado en el principio de localidad:

- 1) Principio de localidad: El evento  $x$  se relaciona causalmente con el evento  $y$ , de modo que el evento  $x$  causa el evento  $y$ :  $F(x)$  es una propiedad causalmente eficaz para la propiedad  $G(y)$ , solo si  $F$  es una propiedad intrínseca de  $x$  (Yoo, 2025).

Lo que el principio de localidad muestra es un punto metafísico acerca de la causalidad que se había indicado anteriormente, esto es que las entidades que interactúan causalmente son aquellas que intervienen en esos mecanismos. Esto se debe al hecho de que factores alejados de la relación causal entre dos entidades que dan lugar a un evento no pueden hacer una *diferencia causal* debido a su naturaleza relacional. Yablo (2003) expresa un punto acerca de este principio del siguiente modo:

Las causas operan a través de mecanismos causales; por lo que deben estar donde los mecanismos están. Específicamente donde se ponen en movimiento (esta idea se remonta a Hume también, en su énfasis en la contigüidad). Dado que el comportamiento emana de mecanismos internos al agente (“la causa de mi acción es aquí y ahora”) sus causas deben ser internas (pp. 316-317).

### 3.2 Eficacia causal

Anteriormente expuse que en un evento causal existen propiedades que serán más o menos relevantes respecto del efecto que causan. Ahora bien, cómo saber qué propiedades son más o menos relevantes para la producción del efecto. Una propuesta se encuentra en Yoo (2025) quien identifica a las propiedades intrínsecas como aquellas que son causalmente eficaces o relevantes para la ocurrencia del efecto. Lo primero que se puede señalar es que la eficacia causal puede remitirse al proceso según el cual ocurre la relación entre entidades, propiedades o eventos en cuestión. Es decir, desde esta perspectiva, si dos eventos “ $x$ ” e “ $y$ ” — asumamos que los relatos son propiedades — se relacionan causalmente de modo que la propiedad  $F(x)$

produce la propiedad  $G(y)$ , esto quiere decir que “ $x$ ” es causalmente eficaz para producir el evento “ $y$ ”, en virtud de la relación intrínseca local entre las propiedades  $F(x)$  y  $G(y)$  respectivamente. Volvamos nuevamente el evento de la botella de vidrio rompiéndose producto de la bala impactando su estructura física en el momento  $t$ . Propiedades intrínsecas como la energía, fuerza, velocidad del evento “bala siendo impulsada por arma”, son propiedades causalmente eficaces respecto del efecto en el evento “la botella de vidrio rompiéndose”. De este modo, la propiedad  $F$  del evento “ $x$ ” es causalmente eficaz respecto de la propiedad  $G$  del evento “ $y$ ”.

Ahora bien, ¿qué ocurre con las propiedades intencionales, es decir, la ocurrencia de un evento mental que causa la conducta? La intuición internista indica que, si un evento mental puede ser eficaz respecto de la conducta que causa, esto ocurrirá sólo por sus propiedades físicas o intrínsecas, estas son las únicas que pueden ser citadas como eficaces para la producción del efecto. En el ejemplo de María yendo al restaurante por un café, las propiedades causalmente eficaces para su comportamiento son propiedades físicas internas o intrínsecas que supervienen en su estructura física. Pero ¿cómo determinar qué propiedades son eficaces? Yoo propone un esquema para la evaluación de esas propiedades. Tal evaluación responde a una práctica bastante extendida en la literatura filosófica, que consiste en introducir contextos modales para evaluar dichas propiedades, es decir, la utilización de contrafácticos. El esquema de Yoo dice algo como esto:

Supóngase que los relata consisten en instancias de propiedades de modo tal que: “ $x$ ” e “ $y$ ” están relacionados causalmente de manera que: “ $x$ ” es el evento de la causa e “ $y$ ” el evento del efecto. A su vez la propiedad  $F(x)$  y  $G(y)$  son respectivamente las propiedades tales por las que “ $x$ ” causa “ $y$ ”; específicamente la propiedad  $G$  es instanciada en “ $y$ ” en virtud del hecho de que “ $x$ ” tiene la propiedad  $F$ .

Entonces la propiedad  $F$  es causalmente eficaz para la propiedad  $G$  si y solo si:

I) La ocurrencia de  $F$  y  $G$  está supeditada causalmente de modo que: Si  $F$  no hubiera ocurrido,  $G$  tampoco lo hubiera hecho.

II) Suponiendo que F y G no han ocurrido, entonces si F hubiera ocurrido, G habría ocurrido.

III) No hay ningún H tal que si H hubiera ocurrido sin F, G no habría ocurrido, o si F hubiera ocurrido sin H, G habría ocurrido de todas formas.

Estas tres condiciones tomadas en su conjunto, indica Yoo, constituyen una prueba para fijar qué propiedades pueden ser catalogadas como eficaces causalmente. Del mismo modo indica que cualquier incumplimiento de alguna de las tres condiciones implicaría que no pasen la prueba. Para visualizar esto en un ejemplo considere el evento de María en la Tierra extendiendo su brazo para tomar un vaso de agua. En este caso, “x” estaría constituido por el deseo de María de tomar agua y su creencia de que el líquido que está en el vaso es agua, y el evento “y” consiste en María extendiendo su brazo para tomar el vaso de agua. Imagínenos ahora una situación contrafáctica, en donde María tiene un vaso de agua al frente de la mesa, y extiende su brazo para coger el vaso y beber agua. En la situación contrafáctica María en el evento “x” ya no instancia la propiedad F, que es ser una creencia acerca de agua — se da de hecho otra creencia, la de ser agua-gemela—. Sin embargo, seguiría siendo el caso que las propiedades relacionadas con la ocurrencia de G se manifiesten en el evento “y”, ya que después de todo, María en la Tierra Gemela toma el vaso para beber agua-gemela. Por lo tanto, se violaría la condición (I), ya que ambas instancian la misma conducta.

Lo que este punto muestra es que, a partir del análisis de contrafácticos, solo las propiedades intrínsecas son causalmente eficaces. Además, extendiendo este análisis al caso de los contenidos de los EM, una consecuencia directa es que solo el contenido que dependa de propiedades intrínsecas será causalmente eficaz, por lo tanto, el contenido externamente individuado no calificaría en dicha prueba.

Un punto importante para recalcar es que existe una relación estrecha desde esta perspectiva entre eficacia causal y relevancia explicativa que fundamentalmente está dada por dos componentes: (I) la estructura causal de la explicación, en efecto, según una larga tradición, explicar un fenómeno consiste en determinar sus causas de modo que la explicación debe ser causal. (II) Dada la naturaleza de la causalidad, y los elementos ya descritos, solo propiedades intrínsecas son causalmente eficaces. Estos dos puntos tomados

en conjunto implican que las propiedades explicativamente relevantes solo pueden ser intrínsecas.

Considero que la afirmación de que solo propiedades intrínsecas o internas son las únicas propiedades relevantes, en términos explicativos, es errónea, ya que las propiedades metafísicas que permiten individuar los factores explicativamente relevantes no se aplican necesariamente a causas o relaciones causales. En las siguientes secciones defenderé que las propiedades relacionales o externas sí cumplen un rol en la explicación de la conducta y mostraré que la asunción internista está equivocada en este punto.

#### **4. Relevancia explicativa**

En la sección anterior introduje el concepto de relevancia causal, en vistas a la eficacia de una propiedad. La idea intuitiva detrás de este concepto es que, en un evento que causa otro evento, existe un conjunto relativamente amplio de propiedades de las cuales unas serán más o menos relevantes que otras para la producción del efecto. Kallestrup (2012) propone la siguiente definición de propiedad causalmente relevante que es ilustrativa: “Una propiedad de una causa responsable de algún efecto, es causalmente relevante si la causa causó ese efecto en virtud de tener esa propiedad” (p. 187). La manera usualmente empleada para comprobar si una causa posee esa propiedad causalmente relevante es, tal como se había indicado, a partir de la implementación de contrafácticos. Es decir, a través de la articulación de escenarios que contengan contextos modales y así determinar si una propiedad pasa o no pasa la prueba. El esquema de Yoo introducido en la sección anterior es un ejemplo de esa prueba.

Lo que el esquema de Yoo muestra es, por un lado, que la implementación de contrafácticos puede servir, al menos a nivel intuitivo, para el análisis de la relevancia explicativa y, por otro lado, que solo propiedades que satisfagan restricciones internistas podrán pasar dicha prueba. Esto se debe a que en esta perspectiva existe una relación muy estrecha entre eficacia causal y relevancia explicativa que está dada por la concepción causal de la explicación. En palabras de Yoo (2025):

El problema es que las propiedades extrínsecas generalmente, como regla general, no pasan la prueba de relevancia causal. Como especificaba la prueba, una propiedad puede ser causalmente relevante sólo si (entre otras cosas) no hubiera ocurrido, el efecto no habría ocurrido; y si hubiera tenido éxito en ocurrir, entonces el efecto se habría producido. Pero este patrón de dependencias contrafácticas no se satisface con contenidos externamente individualizados [...] el gemelo físico idéntico que piensa en XYZ, no en el agua propiamente dicha, hace exactamente lo mismo. Diferentes pensamientos no se manifiestan en diferentes comportamientos. Como resultado, el contenido que contiene estados mentales no es causalmente relevante para el comportamiento (sección 5).

Es decir, el contenido intencional individuado externamente no califica como causalmente eficaz porque no cumpliría con alguna condición contrafáctica y, aparentemente, según la interpretación internista, tampoco puede cumplir un rol explicativo en el ámbito de la conducta, precisamente porque la explicación tiene una estructura causal que requiere eficacia, de la cual estos contenidos carecen. El esquema de Yoo, en este sentido, como criterio de filtro entre propiedades relevantes causalmente, también funciona como un criterio acerca de la relevancia explicativa, dada la relación existente entre eficacia y relevancia explicativa anteriormente indicada.

Sin embargo ¿esto es así? Consideremos dos ejemplos para poner a prueba el criterio.

(I) Supóngase que María va caminando por el centro de la ciudad, con el objetivo de ir hacia la Universidad. Sin embargo, al pronto andar, ve a un grupo de manifestantes y a la policía en un enfrentamiento abierto. De pronto, en ese contexto, una piedra se dirige justo hacia ella, ante lo cual se percata a tiempo y la esquivo moviéndose hacia atrás. Luego María vuelve a su hogar.

(II) Imaginemos que María va caminando por el centro de la ciudad, ya que se dirige hacia la Universidad, cuando en ese momento recibe un llamado de su madre, la cual le dice que su gata está enferma. María, como es una amante de los animales y tiene una relación muy estrecha con su gata, acude de inmediato a su casa.

¿Qué sucede en ambas situaciones? ¿Qué los diferencia en términos de la explicación?  
¿Cómo se comportan los contrafácticos?

El caso (I) parece plausiblemente capturar la idea de que lo que explica la conducta de María puede remitirse a términos puramente intrínsecos. Es decir, lo que explica su conducta de esquivar la piedra y volver a casa se ajusta a las restricciones internistas, a pesar de que lo que gatilla su comportamiento es un estímulo externo. Por ejemplo, María hubiese actuado de la misma manera si la piedra no hubiese existido pero su cerebro hubiese generado accidentalmente el mismo patrón de actividad neuronal, dando lugar a la alucinación de que una piedra vuela en su dirección. Son estos factores intrínsecos los que cumplen con las condiciones contrafácticas relevantes y por lo tanto causales que explican la conducta.

En el caso (II) sin embargo, parece más plausible considerar que la explicación está dada por las propiedades semánticas involucradas. Puesto que podemos mantener todas las propiedades intrínsecas de María estables y, sin embargo, cualquier diferencia contrafáctica en el contenido semántico de la oración “la gata está enferma” determina un curso de comportamiento muy diferente, incluso la no ocurrencia de aquel.

Lo que muestran estos 2 ejemplos es que, por un lado, el esquema de Yoo también puede probar que las propiedades externas o relacionales son relevantes en términos explicativos y, en segundo lugar, que la apelación a contrafácticos funciona tanto para cuentas internistas como para externistas. En este sentido la utilización de contrafácticos como criterio para determinar la relevancia explicativa (i) hacen relevante a las propiedades relacionales y se produce además una interesante tensión ya que, asumiendo el principio de localidad, (ii) recogen propiedades que no son causalmente eficaces.

## **5. Explicación externalista**

El análisis anterior destaca que la utilización de contrafácticos, según se ha descrito, no es suficiente para establecer que solo las propiedades causalmente eficaces son relevantes desde un punto de vista explicativo, especialmente si nos adherimos al principio de localidad en la concepción de causalidad. Este planteamiento sugiere que la relación entre contrafácticos y eficacia causal no debe interpretarse de manera rígida, ya que podría haber contextos explicativos en los que los contrafácticos se articulen en modelos que no necesariamente deben describirse en términos estrictamente causales. En esta sección, mostraré cómo las propiedades externalistas pueden ser genuinamente explicativas, incluso cuando no podemos identificar las propiedades que las convierten en causales.

Para ilustrar este punto podemos considerar cómo funciona la práctica científica asumiendo el principio de localidad. Este principio justificaba una visión acerca de las propiedades que cumplen un rol importante dentro de los contextos causales al señalar que solo las propiedades intrínsecas son causalmente eficaces, porque solo ellas pueden hacer diferencias causales. Si asumimos que las propiedades causalmente eficaces son las explicativamente relevantes, dado los puntos anteriores, entonces solo ellas tienen valor explicativo en el ámbito psicológico.

Refiriéndonos a la práctica científica, existen razones para pensar que esa asunción recién descrita no se cumple en muchas áreas de investigación, en efecto, tal parece ser el caso de disciplinas como la sociología, arqueología o la biología evolutiva. Tal como señala Wilson (1992, p. 115), en este tipo de ciencias parece existir un interés por los procesos y sistemas más que por una entidad aislada. Se tiene en cuenta que esta entidad se inserta dentro de una red explicativa más amplia que involucra diversos factores explicativos de diferente orden, como relaciones y factores históricos del objeto de observación. Podemos visualizar dos ejemplos utilizados por Wilson para ilustrar este punto. El primero de ellos está referido al ámbito de la sociología. Ser criminal es una propiedad que algunas personas instancian pero que no está solamente determinada por hechos físicos acerca de ese individuo (como por ejemplo, alguna predisposición genética), sino que se trata de una propiedad relacional, dado que ser criminal depende de si no se cumplen o se infringen determinadas leyes

estipuladas que rigen el comportamiento de una cierta sociedad. Por lo tanto, la conducta criminal es un tipo de conducta que no superviene en la microestructura local o en las potencias causales intrínsecas únicamente.

El segundo ejemplo se sitúa en el contexto de las ciencias biológicas, específicamente en la biología evolutiva. Wilson sostiene que en las discusiones contemporáneas acerca del término “especie” existe un consenso más o menos generalizado de que la aplicación de este concepto depende tanto de relaciones como variaciones del entorno y no de propiedades intrínsecas que puedan compartir diferentes individuos. Esto quiere decir que, si dos individuos son intrínsecamente indiscernibles aun así pueden pertenecer a especies diferentes. Por lo tanto, las especies se taxonomizan o individualizan en función de propiedades relacionales y no intrínsecas:

[...] La concepción esencialista de las especies, según la cual las especies se definen únicamente en función de las características genéticas intrínsecas o incluso morfológicas de sus miembros, es insuficiente para la explicación en la disciplina.  
[...] La pertenencia de organismo individual no está fijada por las propiedades intrínsecas de ese organismo, sino por las relaciones que tiene con otros individuos.  
[...] Dos organismos podrían ser bioquímicamente idénticos en composición y estructura, y aun pertenecer a especies diferentes —ya que— podrían ser asilados de forma reproductiva y tener filogenias independientes (Wilson, 1992, p. 119).

Hay dos cosas que muestran estos ejemplos: (i) la práctica científica devela que las propiedades explicativas son más amplias que lo que permite la concepción causal de la explicación; (ii) las propiedades externalistas, relacionales como la historia o el contexto, son genuinamente explicativas incluso si no conocemos las propiedades causales. En la siguiente sección, argumentaré que las propiedades externalistas que son explicativamente relevantes no necesitan apelar a la noción de causalidad y explicación antes descrita.

## 5.1 Explicaciones no causales

Qué reflexión podemos hacer respecto de la posición externalista, a partir de lo que se ha expuesto aquí. Por un lado, el internista nos quiere convencer de que se necesita eficacia causal para que una propiedad sea explicativamente relevante, y eso es lo que motiva que el contenido se individua internamente, porque solo de esta manera los contenidos pueden ser eficientes — dado que el contenido de EC no es causalmente eficaz—. Sin embargo, esa motivación pasa por asumir una noción particular de causalidad y explicación. Por otro lado, las estrategias externalistas han consistido en: (I) asumir esa noción de causalidad, pero mostrar cómo las propiedades externalistas pueden tener eficacia causal; (II) permanecer agonísticos sobre cómo las propiedades de EC pueden ser causalmente eficaces, pero mostrar que, independientemente de ese hecho, son genuinamente explicativas. Como se puede observar, la primera estrategia se enfoca más en el lado ontológico, mientras que la segunda en el lado epistémico. Sin embargo, es necesario puntualizar que incluso esta estrategia epistémica no niega que las propiedades de EC sean causales, es solo que ese debate depende de discusiones metafísicas más amplias. En esta sección mostraré que no necesitamos asumir ese compromiso con la causalidad para dar cuenta del valor explicativo de las propiedades externalistas. No obstante, mi objetivo no es una defensa del externalismo, sino más bien aclarar un punto de la discusión.

Tal como se expuso en las secciones anteriores, un modelo influyente de explicación científica tiene una estructura causal en donde el fenómeno a explicar está supeditado bajo una ley. Pero ¿este modelo es lo suficientemente inclusivo para capturar diferentes contextos de explicación? Considere el siguiente enunciado: “el COVID causó una recesión mundial”. Es decir, un virus causó una recesión mundial ¿cómo encajaría ese enunciado en el modelo óntico de la explicación? Explicar la transferencia de un conjunto de trazas a un virus sobre la economía mundial es algo muy extraño. Sin embargo, hay modelos alternativos que nos permitirían capturar este tipo de enunciados.

Para ilustrar este punto podemos tomar el modelo causal de Woodward (2003) que ha sido el modelo más influyente dentro del contexto de las explicaciones en neurociencias. Woodward propone un modelo explicativo cuya idea central es que el mecanismo describe

variables que hacen una diferencia respecto de los valores de otras variables en el modelo y del fenómeno. En este sentido, hacer una diferencia implica la manipulación y la susceptibilidad de intervenir variables. Es decir, en una relación de  $x$  e  $y$ , la variable  $i$ , que es la variable de intervención, puede modificar el valor de una variable, en este caso la variable del efecto, de modo que la intervención hace la diferencia en tanto que puede modificar el valor de la variable. Las condiciones bajo las cuales una variable  $i$  (intervención) puede modificar a una variable  $y$ , en una relación  $x \rightarrow y$ , está dada por satisfacer restricciones contrafácticas. Las condiciones para la intervención de una variable pueden resumirse del siguiente modo:

Sea  $F$  una intervención ideal, entonces:

- (I) Una intervención de  $i$  sobre  $x$  modifica a  $x$  en  $i$ , dado que  $x$  está al alcance de  $i$  en el recorrido de  $i \rightarrow x \rightarrow y$ .
- (II)  $x$  e  $y$ , están en una relación causal si:
  - a)  $i$  no puede ser modificado, a no ser que la modificación ocurra bajo la intervención de  $x$ .
  - b)  $x$  cambia si y solo si  $i$  cambia.
  - c) No puede darse el caso de que una variable  $Z$  sea modificada por  $i$ , y modifique directamente a  $y$ , si no está al alcance de  $i$ .
  - d)  $i$  no debe estar correlacionado probabilísticamente con  $Z$  que modifica  $y$ , por medio del recorrido causal que no va por  $x$ . Estas condiciones tomadas en su conjunto constituyen las relaciones genuinamente causales.

Ahora bien, el propio Woodward (2003) ofrece una definición de explicación no causal:

[...] El elemento común en muchas formas de explicación, tanto causales como no causales, es que deben responder preguntas de “si las cosas hubieran sido diferentes”. Cuando una teoría nos dice cómo cambiaría  $Y$  bajo las intervenciones sobre  $X$ , tenemos (o tenemos el material para construir) una explicación causal. Cuando una teoría o derivación responde a una pregunta de qué pasa “si hubiera sido diferente”,

pero no podemos interpretar esto como una respuesta a una pregunta sobre lo que sucedería bajo una intervención, podemos tener una explicación no causal de algún tipo (p. 221).

Una aplicación del modelo no causal en neurociencias puede encontrarse en Chirimuuta (2017). Lo interesante es que Chirimuuta parte tomando como base las mismas condiciones de suficiencia explicativa que el mecanicismo considera para argumentar su posición, es decir, apelando al modelo manipulacionista de Woodward (2003). Entonces, a partir de relaciones contrafácticas también se puede llegar a una explicación no causal.

Si volvemos al enunciado de que el virus causó una recesión mundial, parece tener más sentido en el modelo de Woodward, ya que podemos entenderlo en el sentido de que el COVID manipuló el mundo de manera que estadísticamente llevó a una recesión. Ese sentido es diferente al sentido óntico de causalidad que tiene en mente Salmon y colegas y que motiva la intuición internista.

Si la intuición respecto de las explicaciones no causales es correcta, esto podría abrir la posibilidad de ampliar el marco explicativo. Consideremos que puede haber explicaciones no causales en los modelos explicativos en neurociencias y se pudiera generalizar el marco explicativo. Quizás pudiera extenderse a otros ámbitos como las explicaciones psicológicas, a partir del análisis de contrafácticos.

## **6. Conclusión**

En este trabajo examiné el argumento causal defendido por algunos filósofos para motivar el internismo. Este argumento establece que las propiedades causalmente eficaces son las únicas propiedades que pueden ser explicativamente relevantes en el ámbito de la conducta y que sólo las propiedades intrínsecas son causalmente eficaces. Dicho argumento se basa en el principio de localidad, pero también en una concepción causal de la explicación ampliamente extendida en la literatura contemporánea sobre la explicación. Argumenté que las propiedades intrínsecas no son las únicas propiedades que son explicativamente

relevantes, porque hay buenas razones para pensar que las propiedades externalistas también lo son en ciertos contextos científicos. Finalmente, y a partir de modelos recientes que exploran la posibilidad de explicaciones no causales, propuse una nueva estrategia que podría ayudar a avanzar en la discusión, ya que las explicaciones no causales dependen de una noción de los contrafácticos, en donde estos no son causales. Esto parece ser compatible con una visión externista del contenido, puesto que a pesar de que el contenido no sea causal, sí parece entrar en relaciones contrafácticas con la conducta. Si eso se puede sostener, entonces podría ser una forma en que las propiedades relacionales tengan relevancia explicativa aun si no poseen eficacia causal.

## Referencias bibliográficas

- Burge, T. (2007). *Foundations of mind*. Oxford University Press.
- Chirimuuta, M. (2018). *Explanation in computational neuroscience: Causal and non-causal*. *British Journal for the Philosophy of Science*, 69(3), 849-880.
- Dretske, F. (1989). Reasons and causes. *Philosophical Perspectives*, 3, 1-15.  
<https://www.jstor.org/stable/2214261>
- Dretske, F. (2015). Supervenience and the causal explanation of behavior. En T. Horgan, M. Sabatés, & D. Sosa (Eds.), *Qualia and mental causation in a physical world: Themes from the philosophy of Jaegwon Kim* (pp. 154-166). Cambridge University Press.
- Galavotti, M. (2005). *Wesley Salmon sobre explicación, probabilidad y racionalidad*, *Enrahonar*, (37), 61-75.
- Hitchcock, C. R. (1995). *Salmon on explanatory relevance*. *Philosophy of Science*, 62(2), 304-320. <https://doi.org/10.1086/289851>
- Kallestrup, J. (2012). *Semantic Externalism*. Routledge.
- Kaplan, D. (1989). Demonstratives. En J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481-563). Oxford University Press.
- Kim, J. (1982). Psychophysical supervenience. *Philosophical Studies*, 41(1), 51-70.  
<https://doi.org/10.1007/BF00353523>
- Kim, J. (1991). Events: Their metaphysics and semantics. *Philosophy and Phenomenological Research*, 51(3), 641-646. <https://doi.org/10.2307/2107887>
- Kim, J. (1993a). Concepts of supervenience. En J. Kim, *Supervenience and mind: Selected philosophical essays* (pp. 53-78). Cambridge University Press. (Trabajo original publicado en 1984).
- Kim, J. (1993b). 'Strong' and 'global' supervenience revisited. En J. Kim, *Supervenience and mind: Selected philosophical essays* (pp. 79-91). Cambridge University Press. (Trabajo original publicado en 1987).

- Kim, J. (1993c). Supervenience for multiple domains. En J. Kim, *Supervenience and mind: Selected philosophical essays* (pp. 109-130). Cambridge University Press. (Trabajo original publicado en 1988).
- Kim, J. (1993d). Supervenience as a philosophical concept. En J. Kim, *Supervenience and mind: Selected philosophical essays* (pp. 131-160). Cambridge University Press. (Trabajo original publicado en 1990).
- Kim, J., & Sosa, E. (1993e). Psychophysical supervenience. In *Supervenience and Mind: Selected Philosophical Essays* (pp. 175-193). Cambridge University Press.
- Kim, J. (1998). *Philosophy of mind*. Westview Press.
- Kripke, S. (1980). *Naming and necessity*. Harvard University Press.
- McGinn, C. (1989). *Mental content*. Blackwell.
- Newman, A. E. (2002). *Causal efficacy and externalist mental content* [Doctoral dissertation, Massachusetts Institute of Technology]. MIT DSpace.  
<https://dspace.mit.edu/handle/1721.1/8147>
- Putnam, H. (1975). *Mind, language and reality: Philosophical papers (Vol. 2)*. Cambridge University Press.
- Robb, D., & Heil, J. (2021). *Metaphysics: A contemporary introduction*. Routledge.
- Ross, L., & Woodward, J. (2022). Irreversible (one-hit) and reversible (sustaining) causation. *Philosophy of Science*, 89(1), 1-10. <https://doi.org/10.1017/psa.2022.70>
- Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.
- Salmon, W. C. (1990). *Four decades of scientific explanation*. University of Minnesota Press.
- Segal, G. (2000). *A slim book about narrow content*. MIT Press.
- Segal, G., & Sober, E. (1991). The causal efficacy of content. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 63(1), 1-30.  
<http://www.jstor.org/stable/4320218>
- Stich, S. (1983). *From folk psychology to cognitive science: The case against belief*. MIT Press.

- Wayne, A. (2022). Explanatory asymmetry in non-causal explanation. *Journal of General Philosophy of Science*, 53, 555-571. <https://doi.org/10.1007/s10838-021-09596-w>
- Wilson, R. A. (1992). Individualism, causal powers, and explanation. *Philosophical Studies*, 68(2), 103-139. <https://doi.org/10.1007/BF00354473>
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
- Yablo, S. (1997). Wide causation. *Philosophical Perspectives*, 11, 251-281.  
<http://www.jstor.org/stable/2216133>
- Yablo, S. (2003). Causal relevance. *Philosophical Issues*, 13, 316-328.  
<http://www.jstor.org/stable/3050536>
- Yoo, J. (2025). *Mental causation*. *Internet Encyclopedia of Philosophy*.  
<https://iep.utm.edu/mental-c/#H5>