



## Explicación e inferencialismo. Una solución antirrealista para la paradoja de la explicación

*(Explanation and inferentialism. An anti-realist solution to the explanation paradox)*

**Resumen:** En este trabajo se propone una posible solución antirrealista a la paradoja de la explicación. Esta propuesta está en línea con la posición instrumentalista de Reiss (2012b, 2013) pero va más allá: se propone una posición inferencialista normativa de la explicación. En primer lugar, describiré las características fundamentales de la paradoja de la explicación (PE), y resumiré las posiciones más importantes al respecto. Luego, mostraré algunos aspectos fundamentales del inferencialismo en el contexto de la explicación, y finalmente, mostraré cómo una posición inferencialista respecto a la paradoja de la explicación se reduce a aceptar las siguientes tres proposiciones: 1. los modelos son verdaderos en el sentido de explicaciones verdaderas, 2. los modelos explican y 3. sólo las explicaciones verdaderas explican. Esto expresa una solución a la paradoja en tanto es perfectamente posible aceptar cada una de las tres proposiciones.

**Palabras clave:** paradoja de la explicación, inferencialismo, explicación, modelos económicos.

**Felipe Agustin Nuñez Michea**  
[fanunezmichea@gmail.com](mailto:fanunezmichea@gmail.com)  
[orcid.org/0000-0003-3571-8146](https://orcid.org/0000-0003-3571-8146)

Profesor de filosofía por la Universidad de Santiago.

Magíster en ciencias económicas por la Universidad de Santiago.

Actualmente estudiante del Research Master in Philosophy and Economics en Erasmus Universiteit Rotterdam.

**Abstract:** In this paper, I propose a possible anti-realist solution to the explanation paradox. This proposal is in line with the instrumentalist position of Reiss (2012b, 2013) but goes further: a normative inferentialist position of explanation is proposed. Firstly, I will describe the fundamental characteristics of the explanation paradox, and summarize the most important positions on the issue. Secondly, I will show some fundamental aspects of inferentialism in the context of explanation, and finally, I will show how an inferentialist position regarding the paradox of explanation reduces to accepting the following three propositions: 1. models are true in the sense of true explanations, 2. models explain, and 3. only true explanations explain. This expresses a solution to the paradox insofar as it is perfectly possible to accept each of the three propositions.

**Keywords:** Explanation Paradox, inferentialism, explanation, economic models.

**Recibido:** 15 de enero de 2024.

**Aceptado:** 25 de junio de 2024.

**Publicado:** 22 de diciembre de 2024.

**Licencia:** CC BY (Felipe Núñez, 2024)



# 1 La paradoja de la explicación y el realismo sobre los modelos

Julian Reiss (2012b; 2013) ha argumentado que existe una paradoja con el concepto de explicación en economía. La paradoja se presenta como un trilema: parece imposible aceptar las tres proposiciones siguientes tomadas en su conjunto, a pesar de que son aceptables, y aceptadas en general, individualmente:

1. Los modelos económicos son falsos
2. Los modelos económicos explican
3. Sólo los relatos (accounts) verdaderos pueden explicar

Este problema se ha denominado "la paradoja de la explicación" (PE) y ha tenido un papel importante en la discusión sobre el rol de los modelos en economía. En este trabajo intentaré proponer una posible solución antirrealista a la paradoja de la explicación que está en línea con la posición instrumentalista de Reiss (2012b; 2013) pero que va más allá: una posición inferencialista normativa de la explicación, una posición que Reiss no defiende pero con la que ha estado trabajando (Reiss, 2012a, 2019). En lo que sigue, en primer lugar, describiré las características fundamentales de la paradoja de la explicación (PE), y resumiré las posiciones más importantes al respecto. Luego, describiré algunas características importantes de la posición inferencialista de la explicación; en tercer lugar, mostraré que desde un punto de vista inferencialista, 2. y 3. son verdaderas, y 1. es falsa, por lo que no hay PE. En este sentido, el inferencialismo es un enfoque posible para dar una solución antirrealista a la PE considerando un concepto unívoco de explicación. En lo que sigue de esta sección, intentaré explicar los movimientos argumentativos básicos que se han sostenido en el debate, en general, cada uno abogando por la negación de al menos uno de los cuernos del trilema.

Algunos autores han negado que los modelos económicos sean falsos. Mäki (2013), Rol (2013) y Hausman (2013), por ejemplo, han argumentado, con algunas diferencias, que los modelos son verdaderos en algunos aspectos importantes, al menos en abstracto.

El argumento general descansa en aceptar el hecho de que los modelos contienen algunas falsedades, pero que, sin embargo, estos no descansan en ellas. Una advertencia al respecto. La frase "Los modelos económicos son falsos" es ambigua en tanto no especifica qué es lo falso de los modelos. Después de todo, solo las proposiciones pueden ser falsas. Hausman (2013), por ejemplo, argumenta que la paradoja descansa en realidad en este tipo de ambigüedades semánticas.

Aquí no pretendo comprometerme con una definición específica de los modelos. Sin embargo, estoy asumiendo que los modelos tienen contenido proposicional, en este sentido, el hecho de que los modelos no descansan en falsedades significa que dentro del conjunto de proposiciones que forman parte de la caracterización de un modelo, las proposiciones falsas no son esenciales. En los siguientes párrafos explicaré a qué me refiero con "proposiciones falsas no esenciales".

Pero antes, consideremos, por ejemplo, el modelo de Schelling (Sugden, 2000) sobre segregación racial espontánea. En este modelo, se busca explicar la existencia

de la segregación a través de las preferencias individuales. En particular, en este modelo se asume que los agentes solo se preocupan por la satisfacción de sus preferencias según las características de sus vecinos: si modelamos agentes con tipos, es decir, con características diferenciadoras compartidas al interior de los tipos, los agentes prefieren vecindarios con personas de su tipo. Así, se explica la segregación en los vecindarios y la tendencia común en las grandes ciudades a la proliferación de barrios de concentración étnica a través de las preferencias individuales, lo que genera el efecto agregado de segregación racial espontánea. Este efecto ocurre incluso si las preferencias por tipo son leves.

Este modelo es una idealización y una falsificación de la realidad en un sentido relevante: no considera precios u otras variables importantes en la decisión de elección habitacional u otros factores que pueden afectar por otros canales que no sean las preferencias individuales.

En este sentido, los modelos son falsos en tanto contienen falsedades. Ahora bien, esto no significa que los modelos descansan en estas falsedades. Después de todo, es perfectamente posible, y bastante probable, que las preferencias individuales sean una causa del fenómeno en cuestión.

Si bien no todos los autores defienden el punto en los mismos términos, la idea de "análisis de robustez derivativo" plateada por Kuorikoski et al., (2010) es ilustrativa: Es posible, en principio, diferenciar en los modelos económicos entre supuestos sustantivos (supuestos que expresan factores causales y sus mecanismos), supuestos galileanos (supuestos sobre la abstracción de otros factores causales no considerados) y supuestos de factibilidad (*tractability assumptions*, supuestos para facilitar los cálculos matemáticos que no deberían, en principio, afectar los resultados del modelo).

El punto central es que si es posible hacer un análisis de robustez, entonces es posible determinar el carácter de las falsedades del modelo: el objetivo fundamental será determinar si las falsedades, esto es, las proposiciones, son sustantivas o galileanas, lo que sería un problema importante en términos de la representatividad del modelo; o más bien de factibilidad, lo que no sería un problema insalvable.

El marco general de argumentación descansa en que hay representaciones verdaderas y representaciones falsas, en eso descansan las suposiciones sustantivas y galileanas. En este sentido, el marco argumentativo es realista: los modelos son representaciones verdaderas de la realidad.

La posibilidad de analizar el carácter de los supuestos se presenta incluso en toy-models muy idealizados. Por ejemplo, en el dilema del prisionero (DP) las funciones de utilidad expresan las preferencias de los jugadores, que se plantean directamente en el conjunto de números expresados en la matriz de pago, mostrando desde el inicio el orden de preferencias en el sentido de preferencias reveladas por acciones hipotéticas. A pesar de su simplicidad, las características de la estructura del juego permiten explicaciones muy útiles (transfieren la comprensión a través de consideraciones contrafácticas) de algunos fenómenos. Por ejemplo, el DP es la estructura básica y la explicación de la necesidad del Leviatán en la filosofía política de Hobbes.

Si queremos comprobar si nuestro modelo se basa en falsedades, o hasta qué punto el modelo descansa de al-

gunos supuestos, podemos realizar un análisis de robustez derivativo. En el caso de la teoría de juegos, un juego no secuencial, como la DP, se define en términos de número de agentes, acciones y funciones de utilidad definidas sobre los perfiles de las acciones. En este caso, es posible diferenciar los supuestos de factibilidad (tractability) de los supuestos básicos. Un supuesto de factibilidad no cambia la estructura del juego en términos de resultados. El número de jugadores no cambia la estructura. Un subconjunto de acciones no cambia el resultado (el conjunto de acciones que dejan intacta la dominancia estricta de "defraudar"). La estructura de la matriz de pagos tiene que mantener la preferencia de perfiles en los jugadores. En estos casos, cada aspecto es independiente. La independencia en este caso se define de forma conceptual: DP es un modelo de teoría de juegos que describe la situación en la que el óptimo de Pareto no es factible porque el único equilibrio de Nash del juego es una estrategia estrictamente dominante.

Un modelo de teoría de juegos es una tupla  $(N, A_i, u_i)$ , donde  $N$  es el número de jugadores,  $A_i$  son las acciones y  $u_i$  es la función de utilidad definida sobre los perfiles de las acciones. Cualquier cambio en  $N$ ,  $A_i$ , o  $u_i$ , es un supuesto de tractabilidad si no cambia la estructura del juego. En este sentido, un supuesto de factibilidad independiente afecta a distintos aspectos de la definición del juego.

Cambiar el juego, por ejemplo, jugándolo de forma secuencial o *ad infinitum* no es un supuesto de factibilidad porque cambia la estructura.

Podemos interpretar el DP como una guerra de precios entre empresas. En ese caso, podemos pensar en diferentes funciones de beneficio para empresas con diferentes, por ejemplo, funciones de costos. Si la estructura del perfil de preferencias se mantiene igual, no hay ningún supuesto central en este sentido.

Por lo tanto, es posible decir en qué sentido este modelo es verdadero: son verdaderas (explicaciones) porque explican por qué la cooperación es inviable y qué caminos pueden cambiar este resultado. Al mismo tiempo, hay una forma de delimitar qué es un supuesto sustantivo y qué no lo es.

En cualquier situación en la que algo pueda considerarse un jugador (una entidad que tiene un conjunto de acciones y preferencias sobre perfiles) y exista una estructura de DP, entonces la DP sería un dispositivo explicativo.

De momento no he ofrecido una definición rigurosa ni de realismo en este contexto ni de explicación. Mi argumento depende de estos conceptos que serán definidos en las secciones siguientes.

## 1.1 Los modelos económicos no explican

Otros autores han negado la segunda proposición: que los modelos económicos expliquen. Alexandova y Northcott (2013), por ejemplo, han argumentado que los modelos tienen muchas características interesantes y útiles en economía, pero explicar no es una de ellas. La principal característica de los modelos es la heurística y la exploración conceptual. Si los modelos explicaran, serían verdaderos y los modelos no son verdaderos. A estos autores no les convence la idea de que sea posible hacer análisis de robustez y diferenciar el carácter de las falsedades. ¿Qué es lo que requiere un modelo para ser verdadero? De esta pregunta depende gran parte de la fuerza argumenta-

tiva de la negación del segundo cuerno del trilema. Más adelante argumentaré que, de hecho, el movimiento argumentativo depende también de la suposición realista, esto es, del marco general que considera que los modelos son representaciones verdaderas de la realidad.

## 1.2 No solo los relatos (accounts) verdaderos pueden explicar

Otros, desde una vereda anti-realista, pero en la misma línea que los realistas que niegan el primer cuerno del trilema, han argumentado que muchos modelos son verdaderos, al menos en abstracto, por lo que es posible un análisis de robustez; pero al mismo tiempo, no todos los modelos ofrecerían explicaciones de hecho (*actual explanations*, explicaciones que describen la estructura causal de los fenómenos), algunos de ellos ofrecen explicaciones de posibilidad (*how-possible explanations*, explicaciones que describen las condiciones bajo las cuales un resultado puede obtenerse, por ejemplo, el teorema de existencia del equilibrio general de Arrow-Debreu, o DP). En este sentido, el concepto de explicación es amplio: explicar no es solo citar causas. Este es el caso de Grüne-Yanoff (2013).

Sugden (2013) ha argumentado, también desde una vereda anti-realista, que no sólo los relatos verdaderos pueden explicar; los modelos explican porque son relatos creíbles en un sentido abductivo: los modelos citan factores causales que son nuestra mejor explicación para algunas regularidades. En este sentido, existe una similitud entre los modelos y el mundo real que está fundamentada en un razonamiento de la mejor explicación según la información disponible.

Sin embargo, para Sugden, la similitud es un concepto explicativo subjetivo, por lo que, en última instancia, la pregunta por el poder explicativo de un modelo no puede tener una respuesta única para la comunidad científica.

## 1.3 PE, realistas y anti-realistas

Puede resultar confuso el modo en que he organizado las posiciones, sobre todo porque, como he dicho, no he definido claramente a qué me refero con realismo y anti-realismo. Siguiendo a Reiss (2012c), los realistas tienen tres características en el contexto de los modelos económicos: i. valoran la ciencia porque la ciencia es la búsqueda de la verdad, ii. los modelos científicos son/deben ser verdaderos en la medida en que representan la estructura causal del mundo (al menos en abstracto), iii. son optimistas sobre nuestra capacidad para encontrar la verdadera estructura causal del mundo.

Una estructura causal determina qué tipo de relaciones son relaciones causales y qué tipo de cosas están implicadas en estas relaciones. Estas cosas se denominan "relatos causales".

Los anti-realistas también tienen tres características: i. valoran la ciencia porque es útil, ii. la ciencia es útil porque hace predicciones sobre la fuerza de las relaciones entre variables, iii. son pesimistas sobre nuestra capacidad para encontrar la verdadera estructura causal del mundo.

Un realista, por ejemplo, podría considerar la distinción entre explicación epistémica y explicación óptica planteada por Salmon (1984). El primer tipo de explicación, la expli-

cación epistémica, no sería una explicación verdadera, sino una cuasi-explicación en la medida en que puede ayudar a disminuir nuestra sorpresa cuando nos enfrentamos a un fenómeno: efectivamente genera algo para quien demanda una explicación. Sin embargo, una verdadera explicación requiere algo más. Esto es justamente lo que ofrece el segundo tipo de explicación, la explicación óptica, ya que esta cita causas, y causas como una característica del mundo. Este camino es exactamente el que toman Alexandrova y Northcott (2013) para argumentar a favor de una solución a PE.

Para Alexandrova y Northcott (2013) solo las explicaciones ópticas son explicaciones verdaderas y esa es una razón importante para descartar el segundo cuerno del trilema. Al mismo tiempo, por eso las explicaciones causales son el caso paradigmático de explicación verdadera (y por eso desde el marco de las explicaciones ópticas, una ausencia nunca puede ser una explicación verdadera).

Lo que buscan los defensores de las explicaciones ópticas es un fundamento metafísico que asegure que sólo los relatos verdaderos puedan explicar. Lo mismo ocurre, en cierto sentido, con los defensores de la primera proposición del trilema: si los modelos son verdaderos es porque los supuestos centrales del mismo son representaciones verdaderas de la estructura causal. En este sentido, sólo los relatos verdaderos explican: sólo los relatos verdaderos citan factores causales verdaderos.

Un anti-realista, como es el caso de Sugden (2013), no puede descansar en la idea de explicación óptica, pues justamente el punto fundamental para el anti-realista es que es posible valorar la ciencia a pesar de no tener modelos verdaderos de la realidad. En este sentido, Sugden niega que solo los relatos verdaderos expliquen: los modelos son más bien explicaciones creíbles útiles, aunque sean explicaciones falsas en última instancia: en muchos sentidos el mercado de autos usados de Arkeloff o el modelo de segregación racial de Schelling, por usar los ejemplos de Sugden (2000), son representaciones falsas de la realidad a la que buscan representar. A pesar de ser explicativos en tanto los factores causales que se describen en los modelos son factores causales plausibles de la realidad representada.

Aquí es donde la propuesta de Sugden me parece problemática. Si bien la propuesta de Sugden no descansa en la idea de explicación óptica, la distinción se cuele en su argumentación, pues acepta que hay explicaciones verdaderas en contraposición a explicaciones falsas. Si no fuese así, ¿por qué hacer la distinción entre relatos verdaderos y relatos falsos, en primer lugar? En cualquier caso, con esto queda claro que para él la verdad no es el único elemento valioso de la ciencia.

Grüne-Yanoff (2013) evade en parte este problema, pues plantea un camino intermedio entre realistas y anti-realistas, defendiendo la idea de que los modelos son verdaderos y que al mismo tiempo explican. Como su punto descansa en la diferencia entre tipos de explicación, permite ampliar el rol de los modelos. Sin embargo, como considera que las explicaciones verdaderas son las explicaciones de hecho (actual explanations), aquellas que citan la estructura causal de los fenómenos, el realismo tiene un papel en su propuesta.

Este es mi punto de partida. Me parece que, justamente, el problema de esta discusión está en el concepto de explicación: la apelación a las explicaciones verdaderas o a las

explicaciones ópticas es errónea o, al menos, es un marco inadecuado para los anti-realistas.

## 2 Representacionalismo e inferencialismo sobre la explicación

En esta sección explicaré algunas características clave del inferencialismo y su relación con la idea de explicación (para una comprensión más amplia del concepto de 'inferencialismo', véase Steinberger, F. & Murzi, J. 2017). Lamentablemente es imposible describir en detalle todos los aspectos implicados en esta discusión. Sin embargo, me limito a discutir todos los aspectos relevantes para la consideración del inferencialismo sobre el concepto de explicación.

Robert Brandom ha planteado (1994) una distinción entre dos formas de entender el contenido conceptual. La primera de ellas es el representacionalismo referencial. En esta tradición el contenido conceptual se entiende como una relación referencial: el fundamento del contenido conceptual es su soporte referencial. Para decirlo con más claridad, lo que demanda el representacionalismo referencial es una causa de nuestros contenidos conceptuales en el concepto de referencia. Esta intuición ha sido extremadamente poderosa en filosofía, y es una parte importante del programa empirista, la idea general de que podemos obtener conocimiento a partir de la experiencia bruta. Si el conocimiento es un tipo de contenido proposicional, la idea es que la experiencia bruta puede darnos contenido proposicional. Con "experiencia bruta" me refiero a un "Ready-made world" en el sentido de Putnam (1982), pero más en general a la idea de que hay algo parecido a datos sensoriales que constituyen la base de nuestro conocimiento.

Sin embargo, hay muchos problemas con esta idea, como demostró Sellars en *Empiricism and Philosophy of Mind* (1956) y el argumento del mito de lo dado (sobre esto, véase DeVries, 2020).

Una segunda forma de entender la idea de contenido conceptual se desarrolla evitando la idea de una teoría de la referencia. El inferencialismo va por esa línea. En él, el contenido conceptual se caracteriza por prácticas normativas de razones y justificaciones. En este caso, el fundamento del contenido conceptual son las relaciones inferenciales.

El inferencialismo intenta mostrar que la pragmática precede a la semántica: la dimensión representacional de la semántica y el contenido conceptual deben entenderse en términos de su papel en el juego de preguntar y dar razones, es decir, una práctica normativa de lo que está permitido en las prácticas discursivas.

Evidentemente, intentar este camino tiene muchas consecuencias de gran envergadura filosófica, sin embargo, este no es el lugar para desarrollarlas. Lo que importa aquí, más bien, es que el inferencialismo está íntimamente relacionado con la explicación.

Según el inferencialismo, nuestras inferencias están garantizadas por el contenido de los conceptos que figuran en las proposiciones. En este sentido, el fundamento de una explicación no es una cuestión metafísica. La cuestión relevante no es qué existe, sino cómo podemos llenar un vacío en la comprensión, que es justamente lo que hacen las ex-



plicaciones, llenar un vacío. Esto coincide con la definición de explicación de Reiss: “Sostengo que una explicación es una transferencia de entendimiento de un *explainer* a un *explainee*, a petición de éste.” (2019, p.38, traducción propia).

Ahora podemos volver a la distinción entre explicaciones epistémicas y explicaciones ónticas (cuyos paradigmas son las explicaciones causales).

Para un inferencialista, o un instrumentalista, la causalidad no es algo relacionado con la metafísica, la causalidad es sólo explicación causal. Reiss, en ese sentido, define la causalidad:

“Para dos agentes distintos en una situación epistémica SE, una afirmación causal que relaciona la causa C y el efecto E es afirmable si y sólo si un agente al citar C en SE explica E causalmente con éxito al otro.” (2019, p. 42, traducción propia).

Para un inferencialista, pero no para un instrumentalista necesariamente, una afirmación causal puede afirmarse si estamos normativamente autorizados por el contenido conceptual de nuestras proposiciones a realizar ese movimiento. Y estamos autorizados en tanto hay una transferencia de entendimiento.

Un pequeño comentario puede ayudar a despejar algunas dudas. Una explicación inferencialista clásica es la teoría deductivo-nomológica de la explicación (DN). DN utiliza inferencias deductivas *formales* para explicar. Godfrey-Smith (2009) resume sus problemas, que por cierto están profundamente asociados a la posición instrumentalista extrema más difundida en economía en el siglo pasado: el instrumentalismo económico de Friedman (1953).

Friedman asume DN como mecanismo de soporte evidencial: una teoría tiene evidencia a favor si y sólo si los hechos son inferencias deductivas de sus hipótesis, es decir, predicciones; y tiene evidencia en contra si y sólo si los hechos no son inferencias deductivas de sus hipótesis.

En este sentido, la centralidad de la predicción en Friedman está en el hecho de que una teoría es un conjunto de enunciados simétricos: no hay razones para plantear que una teoría tiene un punto de partida fundamental (es decir, hipótesis nucleares) ni tampoco para preocuparse por el realismo de las teorías en tanto cumplan con el, a ojos de Friedman, elemento más importante de la ciencia: la predicción.

El inferencialismo que estoy describiendo no es un instrumentalismo *à la* Friedman. El inferencialismo que describo pide inferencias *materiales* del tipo “Si x es coloreado, entonces x tiene extensión” o del tipo “si x es cuadrado, entonces no es redondo”.

En general, diríamos que este tipo de inferencias dependen del contenido de los conceptos. Pero de hecho, estrictamente hablando, es todo lo contrario: el significado de los conceptos depende de las inferencias. Este es justamente todo el punto inferencialista.

Ahora bien, en el uso, lo que podemos inferir de las proposiciones como “X es coloreado” está normado por su propio contenido conceptual.

Si decimos ante un fenómeno que “X es una causa de Y” eso solo cuenta como una explicación si es que efectivamente

en esa situación X tiene las características suficientes para contar como una causa y por ende, genera una transferencia de entendimiento.

Esto nos permite hacer frente a los problemas de simetría en DN y, al mismo tiempo, recoger la importancia de las explicaciones causales: mirar “under the hood” como diría Hausman (1992).

Pero, ¿por qué seguir este camino a contrapelo del empirismo, en cierto sentido? Como he dicho en la primera sección, creo que hay buenas razones para eludir las explicaciones ónticas. Si hay dudas sobre la posibilidad de conocer la verdadera estructura del mundo, o si es inalcanzable (si eres un pesimista sobre el realismo); o si existe la sospecha de que los problemas metafísicos son sólo problemas de lenguaje; o que ninguna de las teorías actuales de la causalidad es satisfactoria en un sentido de equilibrio reflexivo; entonces ser un inferencialista, al menos metodológicamente, o un instrumentalista, es una posible salida. En cualquier caso, el objetivo de este trabajo no es ofrecer una defensa del inferencialismo en general, es sólo mostrar cómo un relato inferencialista puede resolver la paradoja descartando la primera proposición de PE.

### 3 Inferencialismo y PE

En esta sección, explicaré cómo el inferencialismo trata con PE.

#### 3.1 Los modelos explican (2)

Alexandrova y Northcott (2013) consideran que los modelos tienen muchas características interesantes y útiles en economía, pero explicar no es una de ellas. En economía, argumentan, los modelos no explican porque, por un lado, la principal característica de los modelos es la heurística y la exploración conceptual, no la explicación. Y por otro lado, defienden la teoría óntica de la explicación.

Si bien es perfectamente posible y razonable el papel que Alexandrova y Northcott atribuyen a los modelos, no hay una contradicción entre heurística y explicación. En este sentido, habría que determinar: 1. por qué razón la explicación no puede ser un papel y 2. cómo explicar por qué, pese a la opinión de Alexandrova y Northcott, la gran masa de economistas considera que los modelos sí son explicativos.

En cualquier caso, este último punto se aleja de los objetivos de este trabajo.

El punto fundamental se centra en el primer aspecto. Es que, justamente, Alexandrova y Northcott solo pueden defender que la explicación no juega ningún papel en los modelos económicos a condición de aceptar una perspectiva óntica de la explicación. Pero, como he mostrado en la sección anterior, es posible defender explicaciones epistémicas en términos del inferencialismo. En ese sentido, es perfectamente posible aceptar que los modelos explican.

Por supuesto, alguien puede preguntar qué hace que una inferencia tenga éxito. Solo se puede ofrecer una respuesta: la estructura y las relaciones de las proposiciones relacionadas con la transferencia de comprensión. Si esto es insuficiente, la discusión es, en cualquier caso, mucho más amplia.

Alguien podría argumentar que si bien no hay problemas con la definición de explicación ofrecida por el inferencialismo, no es cierto que los modelos, al menos los modelos económicos, ofrezcan un mayor entendimiento. Esta defensa debe demostrar que los modelos económicos no ofrecen un mayor entendimiento de los fenómenos económicos para los propios economistas. Esta es una posición difícil de creer.

En cualquier caso, es interesante profundizar en este aspecto, ya que implica tener cierta claridad sobre cuáles son los fenómenos económicos (esto es, por ejemplo, tener un conjunto definido de factores causales) y sobre los aspectos centrales de la teoría económica (¿es acaso la teoría microeconómica? ¿La teoría del equilibrio general? ¿O más bien la revolución en economía empírica y aplicada?), lo que es justamente un problema si se considera que la ciencia económica es una ‘ciencia de modelos’.

### 3.2 Sólo los relatos verdaderos pueden explicar (3)

Es engañoso considerar la tercera proposición del trilema de la forma en que se ha escrito. Pueden plantearse varias cuestiones: ¿Qué se considera un relato? ¿Es un relato una narración que describe y explica? Es muy diferente una descripción verdadera que ser una explicación verdadera, después de todo. O al menos es posible plantear una diferencia entre ambas si somos neutrales sobre lo que significa explicación en términos de no aceptar de antemano una perspectiva óptica.

Por ejemplo, es posible decir que una descripción verdadera necesita referirse a los elementos verdaderos citados en la descripción. Hay razones para pensar que no es necesario en la explicación aceptar de antemano la perspectiva óptica. Solo un ejemplo: la ley de gases ideales no se refiere a los elementos verdaderos que desempeñan un papel en la ontología de la física, pero es defendible como explicación de algunos fenómenos a nivel macro.

Si es posible tener una explicación sin descripción verdadera, esto nos lleva a una nueva formulación de 3’:

3’: Sólo las explicaciones verdaderas explican

Esta es una proposición mal formulada. Me sorprendería que alguien sostuviera que las explicaciones falsas explican. Si alguien piensa que las explicaciones están relacionadas con descripciones verdaderas, entonces las explicaciones falsas no explican. Si alguien piensa que las explicaciones deben reducir la sorpresa o que la explicación es una transferencia de comprensión, entonces una explicación verdadera se define por lo que las explicaciones hacen a los agentes, no por las características del mundo: en este sentido, las dos últimas aproximaciones son epistemológicas en el sentido de Salmon.

El punto fundamental es que si 3’ es cierta de forma no trivial, entonces supone que la verdad es sobre otro tipo de cosas, no sobre proposiciones. Seguramente, sobre qué tipo de cosas existen y pueden funcionar como relatos causales. En este sentido, 3’ asumiría un enfoque óptico de la explicación.

En cambio, si consideramos 3’ desde un punto de vista

inferencialista, es trivialmente cierta. Si una explicación es una transferencia de comprensión tras una petición, entonces en cualquier caso en el que haya una transferencia de comprensión hay una explicación verdadera.

Este es, evidentemente, un enfoque deflacionista de la verdad en la explicación: la explicación no necesita una descripción verdadera de la realidad (en términos de relatos causales), solo necesita esa transferencia de comprensión para tener éxito. En este sentido, las explicaciones verdaderas explican.

### 3.3 Los modelos son verdaderos

Dada la discusión anterior, está claro que, desde un punto de vista inferencialista, los modelos son verdaderos al menos en el sentido de explicaciones verdaderas. Como hemos visto en la primera parte, los críticos aceptan que los modelos explican, excepto Alexandrova y Northcott (2013) que aceptan explícitamente que solo las explicaciones ópticas son explicaciones verdaderas. En cualquier caso, si Alexandrova y Northcott aceptasen que las explicaciones epistémicas son explicaciones verdaderas, entonces los modelos explicarían justamente por su poder heurístico.

Ahora bien, si los modelos son algo más que explicaciones, habría que defenderlo con otro argumento de espíritu realista. Esta discusión ha mostrado que si los modelos son explicaciones, los modelos son verdaderos en ese sentido. Dado lo anterior, se sigue que no hay PE tal y como ha sido planteada en la introducción. Si se considera que no es necesario aceptar una noción de explicación óptica y se acepta que los modelos explican, entonces es cierto que:

1. Los modelos son verdaderos en el sentido de explicaciones
2. Los modelos explican
3. Sólo las explicaciones verdaderas explican

Lo anterior muestra que la paradoja depende de aceptar cierto realismo implícito y que sin él, no hay paradoja alguna.

Es interesante cómo se relaciona esta solución con otras perspectivas recientes. Rohwer and Rice (2016) han intentado, en una discusión más general, mostrar que los modelos pueden ser, por una parte, explicativos, y por otra, pueden ser útiles en tanto ayudan al entendimiento de ciertos mecanismos explicativos. Si los modelos son explicativos o parcialmente explicativos, los modelos tendrían una “relación metafísica” con la explicación; si los modelos solo son útiles en tanto entendimiento, tendrían una relación “epistemológica”. La distinción asume que una explicación verdadera debe citar elementos verdaderos de la realidad, o al menos algunos de ellos, en el caso de las explicaciones parciales; y por otro lado, que una relación epistemológica entre modelo y explicación es un avance en el entendimiento en el sentido de exploraciones conceptuales o heurística, en la línea de Alexandrova y Northcott. Sin duda esto les permite tener una visión más pluralista respecto a la utilidad de los modelos, sin embargo, sigue manteniéndose el problema esencial: se asume una noción óptica de la explicación y se asume que el entendimiento no es una parte constituyente de la explicación misma, no

<sup>1</sup>Agradezco al revisor anónimo que levantó este punto.

un elemento auxiliar. En el relato inferencialista que he ofrecido estas relaciones quedan establecidas.

Verreault-Julien (2019)<sup>1</sup>, también se acerca por esta línea, mostrando explícitamente que la idea de entendimiento no depende de las explicaciones causales, abriendo paso a posiciones anti-realistas. Van Riel (2017) ofrece un análisis lógico de la paradoja mostrando que el elemento más interesante de la misma se encuentra en un análisis epistemológico: un análisis que toma en cuenta la relación entre entendimiento y explicación. Según Van Riel, la paradoja surge genuinamente porque es imposible aceptar que un sujeto S crea que un modelo M ofrece entendimiento sobre un fenómeno X a través de Y y al mismo tiempo S acepte que M es falso explicativamente (es decir, que no es cierto que X porque Y).

Creo que Van Riel está en lo correcto. El relato que he desarrollado aquí es consistente con el punto: los modelos no son falsos en el sentido de explicaciones falsas. Justamente esto implica tener una clara idea de qué significa explicar y ese es el punto que ilumina un enfoque inferencialista.

## 4 Observaciones finales

He ofrecido una manera de tratar el PE en un marco inferencialista. He mostrado que, dado este marco, hay buenas razones para aceptar 2 y 3 y rechazar 1.

En la introducción he dicho que la apelación a explicaciones verdaderas o explicaciones ónticas es errónea o, al menos, es un marco inadecuado para los anti-realistas. Ahora estoy en condiciones de defenderlo.

En el caso de Grüne-Yanoff (2013), aunque acepta que hay distintos tipos de explicaciones, no ofrece una idea general de lo que constituye una explicación en general ni de por qué solo las explicaciones reales son explicaciones verdaderas.

En el caso de Sugden (2013), considera que los modelos son falsos pero creíbles y la credibilidad se basa en la similitud, que es un concepto explicativo subjetivo. Para Sugden, el poder explicativo es una cuestión subjetiva y cada economista tiene apreciaciones diferentes sobre la similitud de los modelos con la realidad. Esto es un problema metodológico importante porque significaría que no hay una forma objetiva de evaluar los modelos económicos.

Un enfoque inferencialista, por otro lado, nos permite dar una definición a una explicación que se basa en la comprensión, que a su vez está relacionada con el conocimiento. Me resulta muy difícil aquí dar una explicación exhaustiva de cómo creo que se relacionan la comprensión y el conocimiento, pero creo que es muy plausible considerar que la comprensión requiere, al menos, verdad en un sentido inferencialista: inferencias materiales consistentes entre entidades y relaciones asumidas por los modelos.

Al mismo tiempo, y esto se sigue de la última frase, el inferencialismo es consistente con la evaluación de teorías; inferencialismo no implica *anything goes*: es perfectamente posible entender cómo funciona lo que las teorías dicen que existe (en términos de relaciones inferenciales materiales) y cómo cumplen las condiciones de relaciones causales (por ejemplo, desde un marco manipulabilista).

Ésta puede ser una forma no estándar de interpretar diferentes supuestos en el análisis de robustez (Kuorikoski,

Lehtinen, and Marchionni, 2010), pero esto queda para una investigación posterior.

## 5 Agradecimientos

Agradezco a FONDECYT y a los miembros del FONDECYT N 1241630 sin los cuales este trabajo no podría haber sido posible. Agradezco también a los miembros de Erasmus Institute for Philosophy and Economics (EIPE), de Erasmus University Rotterdam. Finalmente agradezco a los dos revisores anónimos por sus valiosos comentarios y sugerencias.

## References

- Alexandrova, A. and Northcott, R. (2013). It's just a feeling: why economic models do not explain. *Journal of Economic Methodology*, 20(3):262–267. doi:10.1080/1350178x.2013.828873.
- Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard University Press.
- deVries, W. (2020). Wilfrid Sellars. In Zalta, E. N. and Nodelman, U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2020 edition. URL.
- Friedman, M. (1953). *Essays in positive economics*. University of Chicago Press.
- Godfrey-Smith, P. (2009). *Theory and reality: An introduction to the philosophy of science*. University of Chicago Press.
- Grüne-Yanoff, T. (2013). Genuineness resolved: a reply to reiss' purported paradox. *Journal of Economic Methodology*, 20(3):255–261. doi:10.1080/1350178x.2013.828866.
- Hausman, D. M. (1992). *Essays on philosophy and economic methodology*. Cambridge University Press. doi:10.1017/cbo9781139173964.
- Hausman, D. M. (2013). Paradox postponed. *Journal of Economic Methodology*, 20(3):250–254. doi:10.1080/1350178x.2013.828868.
- Kuorikoski, J., Lehtinen, A., and Marchionni, C. (2010). Economic modelling as robustness analysis. *The British Journal for the Philosophy of Science*. doi:10.1093/bjps/axp049.
- Mäki, U. (2013). On a paradox of truth, or how not to obscure the issue of whether explanatory models can be true. *Journal of Economic Methodology*, 20(3):268–279. doi:10.1080/1350178x.2013.828869.
- Putnam, H. (1982). Why there isn't a ready-made world. *Synthese*, pages 141–167. doi:10.1007/bf00413825.
- Reiss, J. (2012a). Causation in the sciences: An inferentialist account. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(4):769–777. doi:10.1016/j.shpsc.2012.05.005.
- Reiss, J. (2012b). The explanation paradox. *Journal of Economic Methodology*, 19(1):43–62. doi:10.1080/1350178x.2012.661069.

- Reiss, J. (2012c). Idealization and the aims of economics: three cheers for instrumentalism. *Economics & Philosophy*, 28(3):363–383. doi:10.1017/s0266267112000284.
- Reiss, J. (2013). The explanation paradox redux. *Journal of Economic Methodology*, 20(3):280–292. doi:10.1080/1350178x.2013.828874.
- Reiss, J. (2019). Causal explanation is all there is to causation. *Teorema: Revista Internacional de Filosofía*, 38(3):25–52.
- Rohwer, Y. and Rice, C. (2016). How are models and explanations related? *Erkenntnis*, 81:1127–1148. doi:10.1007/s10670-015-9788-0.
- Rol, M. (2013). Reply to julian reiss. *Journal of Economic Methodology*, 20(3):244–249. doi:10.1080/1350178x.2013.828870.
- Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press. doi:10.1515/9780691221489.
- Sellars, W. (1956). Empiricism and the philosophy of mind. *Minnesota studies in the philosophy of science*, 1(19):253–329.
- Steinberger, F. and Murzi, J. (2017). Inferentialism. In *Blackwell Companion to Philosophy of Language*, pages 197–224. Wiley Blackwell. doi:10.1002/9781118972090.ch9.
- Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology*, 7(1):1–31. doi:10.1017/cbo9780511819025.032.
- Sugden, R. (2013). How fictional accounts can explain. *Journal of Economic Methodology*, 20(3):237–243. doi:10.1080/1350178x.2013.828872.
- Van Riel, R. (2017). What is the problem of explanation and modeling? *Acta Analytica*, pages 263–275. doi:10.1007/s12136-016-0307-y.
- Verreault-Julien, P. (2019). Understanding does not depend on (causal) explanation. *European Journal for Philosophy of Science*, 9(2):18. doi:10.1007/s13194-018-0240-6.