

rcfc

REVISTA COLOMBIANA DE FILOSOFÍA DE LA CIENCIA

2024 enero - junio

ISSN: 0124-4620



UNIVERSIDAD
EL BOSQUE

Departamento de Humanidades
Programa de Filosofía
Maestría en Filosofía de la Ciencia

rcfc

REVISTA COLOMBIANA DE FILOSOFÍA DE LA CIENCIA

Indexada en
Philosopher's Index
Red de revistas científicas de América Latina y el Caribe, España y Portugal (RedALyC)
Emerging Sources Citation Index (ESCI)
Red Iberoamericana de Innovación y Conocimiento Científico (REDIB)
Directory of Open Access Journals (DOAJ)
Dialnet



UNIVERSIDAD
EL BOSQUE

Departamento de Humanidades
Programa de Filosofía
Maestría en Filosofía de la Ciencia

© Revista Colombiana de Filosofía de la Ciencia
ISSN: 0124-4620 ISSN: 2463-1159. <https://doi.org/10.18270/rcfc>
Volumen xxiv No. 48
2024 enero – junio

Editor Edgar Eslava

Editor Asistente José Álvarez

Asistente Editorial Andrés Felipe Montañez

Comité Editorial Gustavo Caponi, *Universidad Federal de Santa Catarina*. Flor Emilce Cely, *Universidad El Bosque*. William Duica, *Universidad Nacional de Colombia*. Olimpia Lombardi, *Universidad de Buenos Aires*. Angela Bejarano, *Phd Universidad Pedagógica nacional, Colombia*. Felipe Bravo, *PHD Investigador Independiente, Colombia*

Comité Científico Rafael Alemañ, *Universidad Miguel Hernández, España*. Alfredo Marcos, *Universidad de Valladolid, España*. Nicholas Rescher, *Universidad de Pittsburg, EE.UU*, Ivana Anton Mlinar, *Universidad de Cuyo, Argentina*.

Fundador Carlos Eduardo Maldonado, Universidad El Bosque

UNIVERSIDAD EL BOSQUE

Rectora María Clara Rangel Galvis, OD, MSC, PHD

Vicerrectora Académica Natalia Ruiz Rodgers PhD

Vicerrector Administrativo Francisco José Falla Carrasco

Vicerrectora de Investigaciones Gustavo Silva Carrero

Director del Departamento de Humanidades Camilo Duque Naranjo

Directora del Programa de Filosofía Ana Isabel Mendieta P.

Director de la Maestría en Filosofía de la Ciencia José Luis Cárdenas PhD.

Corrección de estilo Daniela Echeverry Ortiz danyohtarwen@gmail.com

Traducción José Álvarez & María Paula Parra m92parra@gmail.com

Concepto, diseño, diagramación y cubierta Editorial Universidad El Bosque

Solicitud de canje Universidad El Bosque, Biblioteca – Canje, Bogotá - Cundinamarca - Colombia, biblioteca@unbosque.edu.co

Correspondencia e información Universidad El Bosque, Departamento de Humanidades, Cra. 7B # 132-11, Tel. (57-1) 258 81 48, revistafilosofiaciencia@unbosque.edu.co



CONTENIDO

La economía <i>folk</i> y la ciencia económica	9
<i>Luis Enrique Ortiz Gutiérrez - México</i>	
Traducción: Folk economics and economic science	45
<i>Luis Enrique Ortiz Gutiérrez - México</i>	
Elementos para una revisión epistemológica de la axiomática de Kolmogorov	79
<i>Alberto Landro & Mirta L. Gonzalez - Argentina</i>	
Traducción: Elements for an epistemological review of Kolmogorov axiomatics	109
<i>Alberto Landro & Mirta L. Gonzalez - Argentina</i>	
El simio egoísta: críticas a la hipótesis del pensamiento maquiavélico	137
<i>Mateo Arias-Vélez - México</i>	
La posibilidad de explicación científica a partir de modelos basados en redes neuronales artificiales	161
<i>Alejandro E. Rodríguez-Sánchez - México</i>	
Entidades físicas, identidad temporal y basura espaciotemporal. Una reconstrucción crítica de la ontología física de Dowe	195
<i>Wilfredo Quezada Pulido & Luis Pavez - Chile</i>	
Holismo local y cambio semántico en la teoría de Kuhn: Aristóteles, Newton y la inconmensurabilidad taxonómica	231
<i>Daniel Labrador-Montero - España</i>	
Indicaciones para los Autores	275

CONTENTS

Folk economics and economic science	9
<i>Luis Enrique Ortiz Gutiérrez - México</i>	
Translation: Folk economics and economic science	45
<i>Luis Enrique Ortiz Gutiérrez - México</i>	
Elements for an epistemological review of Kolmogorov axiomatics	79
<i>Alberto Landro & Mirta L. Gonzalez - Argentina</i>	
Translation: Elements for an epistemological review of Kolmogorov axiomatics	109
<i>Alberto Landro & Mirta L. Gonzalez - Argentina</i>	
The selfish ape: critics to hypothesis of machiavellian thinking	137
<i>Mateo Arias-Vélez - México</i>	
The possibility of scientific explanation from models based on artificial neural networks	161
<i>Alejandro E. Rodríguez-Sánchez - México</i>	
Physical entities and spatiotemporal junk. A critical reconstruction of p. Dowe's physical ontology	195
<i>Wilfredo Quezada Pulido & Luis Pavez - Chile</i>	
Local holism and semantic change in the kuhn's theory: Aristotle, Newton and the taxonomic inconmensurability	231
<i>Daniel Labrador-Montero - España</i>	
Instructions for Authors	279

LA ECONOMÍA *FOLK* Y LA CIENCIA ECONÓMICA*

FOLK ECONOMICS AND ECONOMIC SCIENCE

LUIS ENRIQUE ORTIZ GUTIÉRREZ
Universidad de Guadalajara
Guadalajara, jalisco.

Luis.Ogutierrez@academicos.Udg.Mx
<https://orcid.org/0000-0003-2680-9230>



RESUMEN

El desarrollo de la ciencia ha producido una brecha entre el conocimiento aceptado por las comunidades científicas y las creencias populares. Esto ha preocupado a los científicos al considerar que muchas de esas creencias resultan ser insostenibles. Además de los problemas relativos a la educación y la divulgación científica, el estudio de las creencias *folk* plantea interesantes interrogantes para la epistemología y las ciencias cognitivas acerca de los procesos de formación y justificación del conocimiento. Este tema tiene también importantes repercusiones en cuanto a las acciones, partiendo de que las creencias *folk* pueden motivar decisiones erróneas. El objetivo del presente artículo es analizar las creencias *folk* relativas al ámbito económico, mostrando algunas de sus posibles causas y algunas consecuencias en lo tocante a la toma de decisiones.

Palabras clave: epistemología; conocimiento; creencia; creencia *folk*; ciencia *folk*; economía; ventaja comparativa epistémica; sesgos cognitivos; moral; política.

* Este artículo se debe citar: Ortiz Gutiérrez, Luis Enrique. "La economía *folk* y la ciencia económica". *Revista Colombiana de Filosofía de la Ciencia* 24.48 (2024): 9-44. <https://doi.org/10.18270/rcfc.4229>

ABSTRACT

The development of science has produced a gap between the knowledge accepted by scientific communities and popular beliefs. This subject has worried scientists, considering that many of these beliefs turn out to be untenable. In addition to the problems related to education and scientific dissemination, the study of folk beliefs raises intriguing questions for epistemology and cognitive sciences about the processes of formation and justification of knowledge. This theme also has important repercussions in terms of actions, based on the fact that folk beliefs can motivate wrong decisions. The objective of this article is to analyze the folk beliefs related to the economic field, showing some of its possible causes and some consequences regarding decision making.

Keywords: epistemology; knowledge; belief; folk belief; folk science; economics; epistemic comparative advantage; cognitive bias; morals; politics.

“The feeling of believing is a more or less sure indication of there being established in our nature some habit which will determine our actions.

Doubt never has such an effect”

Charles Sanders Peirce (1991 149).

1. LA BRECHA ENTRE EL CONOCIMIENTO CIENTÍFICO Y LAS “CIENCIAS *FOLK*”

Uno de los problemas teóricos que ha ocupado a especialistas de diversas disciplinas es la brecha que se ha dado entre las creencias populares y el conocimiento científico. Por una parte, existen diversas creencias, adoptadas por diversos sectores de la población, acerca de cómo es el universo, los seres vivos, la sociedad, el lenguaje, la mente, etc., que son transmitidas por generaciones. Por otro lado, los científicos han formu-

lado diversas teorías, sometidas a diferentes pruebas experimentales o razonamientos deductivos, para intentar explicar estos objetos o fenómenos, lo que ha llevado a las ciencias a distanciarse progresivamente de estas creencias (Wolpert 1992). En este sentido, la brecha ha crecido en la medida en que los resultados de la investigación científica contradicen o cuestionan dichas creencias.

Podemos mencionar algunos ejemplos: las personas sin formación en medicina tienen múltiples creencias acerca de las enfermedades que no concuerdan con lo que plantean los médicos: a partir de ciertos síntomas que manifiesta una persona, las personas intuitivamente los asocian con una enfermedad determinada, por ejemplo, el resfriado, aunque estos mismos síntomas se pueden manifestar en diferentes padecimientos, que van desde la gripe común hasta los distintos tipos de influenza. Así, los legos en medicina suponen que ciertos remedios que “curan” el resfriado servirán en todos esos casos —por ejemplo, el caldo de pollo—, cuando las infecciones respiratorias virales no tienen una “cura” en realidad, sino que los médicos prescriben a los pacientes fármacos analgésicos, antihistamínicos y antieméticos para sobrellevar los síntomas.

Otro ejemplo lo representan las concepciones lingüísticas legas. El lingüista Ray Jackendoff (2003 3-4) enlista una serie de creencias acerca de la lengua: 1) “Hay una forma correcta de hablar una lengua” (en contraste, los lingüistas rechazan que exista un “uso correcto” de la lengua y señalan que los múltiples usos de la lengua se deben a variaciones dialectales, sociolectales, diacrónicas, etc.); 2) “La lengua se está deteriorando por los malos usos” (en esta creencia, se ignora el cambio lingüístico tanto en el léxico como en la gramática, el cual es inevitable); 3) “Los diccionarios nos informan sobre la forma correcta de hablar” (pero los diccionarios no son más que inventarios, incompletos y parciales, de ciertas muestras léxicas, muchas de ellas pertenecientes a variantes dialectales); 4) “La *lengua escrita* es la lengua” (pero, históricamente, la aparición de la escritura es tardía y muchas lenguas del presente no cuentan con un sistema de escritura propio). En términos generales, estas creencias de los hablantes asumen que hay un modelo normativo rígido de la lengua y cualquier uso que no se corresponda con este modelo será considerado “incorrecto”.

De este modo, existe una gran variedad de creencias acerca de la naturaleza, la salud, la lengua, la vida, la sociedad, etc., que contrastan con los principios acepta-

dos por los científicos. Desde luego, la existencia de tales creencias no representa un problema. Como señalaba el filósofo norteamericano Charles S. Peirce (1991 144-ss), las creencias representan un punto intermedio entre la duda y el conocimiento científico. Pero las creencias no son únicamente conocimientos provisionales con los cuales los sujetos conforman una visión del mundo que les genera un cierto grado de certeza, sino que además estos producen motivaciones para la acción. Permanecer en el estado de la duda mantiene al sujeto en la inacción y, por ello, requiere de cierta certeza que le otorga la creencia para actuar en el mundo.

En todo caso, el distanciamiento entre las creencias populares y los enunciados científicos ha aumentado significativamente. Es posible que tal distanciamiento obedezca a ciertas divergencias entre los procesos de formación de conocimiento: mientras las teorías científicas dependen de complejos procesos de elaboración cognitiva, los sistemas de creencias populares parecen ser producidos por intuiciones de los sujetos, por tradiciones culturales, entre otros posibles factores. Esto significa que la comprensión de las teorías científicas requiere de una formación teórica y una especialización cada vez mayor, lo que podría explicar que la brecha se expanda aún más.

Desde hace algunos años, este tema ha sido desarrollado en diferentes estudios en el interior de algunas disciplinas científicas. En psicología cognitiva, por ejemplo, Steven Pinker (2002 195-196) presenta un listado de ciertas “facultades cognitivas” que permiten al ser humano formar conocimientos intuitivos con los cuales intenta comprender y, en consecuencia, actuar en el mundo circundante. Este listado, que el autor reconoce como provisional, incluye una *física intuitiva*, una *biología intuitiva*, una *lingüística intuitiva*, una *ingeniería intuitiva*, una *psicología intuitiva*, una *economía intuitiva*, etc.

Otras investigaciones han ahondado más en el fenómeno. Bajo el adjetivo ‘folk’ se han agrupado las creencias populares en torno a algunos objetos de estudio propios de ciertas disciplinas: se habla, por tanto, de “física folk” (Hayes 1979, 1983; Smith 1994), “biología folk” (Brown 1977, 1979), “psicología folk” (Pinker 2002), “lingüística folk” (Preston 2006), etc. La finalidad de estos estudios es identificar las creencias folk y contrastarlas con las afirmaciones científicas para mostrar sus diferencias. Algunos de los propósitos son intentar desarrollar estrategias para la divulgación

de las ciencias, la enseñanza del conocimiento científico y el asesoramiento para las legislaciones en materia de ciencia y tecnología, entre otros.

La pretensión del presente escrito es analizar las creencias *folk* en una disciplina muy particular: la economía. En especial, se enfocará en identificar algunas de las creencias económicas *folk* más comunes y presentar algunas hipótesis acerca de sus posibles causas. Desde luego, tales hipótesis requerirán de una extensa corroboración empírica, por lo cual este trabajo representa el esbozo de un programa de investigación que puede tener implicaciones importantes para la epistemología y la filosofía de la ciencia, por no mencionar la posible formulación de algunas directrices para el desarrollo de políticas educativas tendientes a la promoción y divulgación del conocimiento científico.

Más adelante, revisaremos algunas implicaciones que podría tener la economía *folk* para la política. Si las creencias, como sostiene Peirce, establecen motivaciones para la acción, entonces su estudio puede ser de capital importancia para describir y explicar las conductas de los individuos. Esto adquiere mucha relevancia para las democracias modernas, en tanto que las políticas públicas son decididas a partir de la participación de los votantes en los procesos electorales. Mas si los votantes toman decisiones políticas partiendo de creencias equivocadas, ¿qué consecuencias podrían tener para la democracia?

2. DEFINICIONES

En epistemología, los filósofos han discutido largamente acerca de la relación entre el conocimiento y las creencias. Esto ha conducido a reflexionar de qué forma se deben definir estas nociones de forma más clara y precisa. Desde luego, como suele ocurrir en filosofía, todas las definiciones son problemáticas sin excepción y lo que se presentará a continuación no está exento de ello.

En primera instancia, es preciso aclarar que la distinción entre *creencia* y *conocimiento* o entre *público experto* y *público lego* obedece simplemente a la finalidad de distinguir ciertos grupos de individuos y determinados tipos de conocimiento. Hablar

de *público lego* no significa, de ningún modo, caracterizar a ciertas personas como ‘ignorantes’ o ‘personas con poca capacidad intelectual’, ya que, de alguna forma, *todos somos legos*: es imposible tener un conocimiento pleno de toda disciplina teórica, y, en realidad, los seres humanos se especializan en ciertos conocimientos, sean teóricos o prácticos, según sus capacidades, preferencias o intereses individuales. Así, un matemático puede ser un experto reconocido en su área, pero desconocer la biología básica, y un biólogo puede ser una autoridad en genética de poblaciones, pero incompetente para cambiar la llanta de su automóvil. Esta aclaración es importante para evitar que el uso de estos términos produzca confusiones, malentendidos o equívocos.

Para efectos de la presente exposición, se adoptará la definición de creencia, propuesta por el filósofo H. H. Price (1935 232) —la cual es una reformulación de la definición del también filósofo Cook Wilson—. Esta establece que una creencia consiste en lo siguiente:

Dado un sujeto S, se dice que S cree en la proposición p
si S tiene presente o considera p
y S asiente o adopta p.

Para ilustrar esta definición, supongamos que Bety ignora dónde se halla su gato. Bety considera o tiene presente una serie de proposiciones, “el gato se halla en la alacena” (A), “el gato se encuentra en el comedor” (C) y “el gato está en el jardín (J)”, como posibles descripciones acerca de la ubicación de su mascota felina. Si Bety asiente o adopta C, descartando con ello a A y J, entonces se puede afirmar que “Bety cree que C” (“Bety cree que el gato se encuentra en el comedor”), esto es, Bety asume que C es verdadera, con lo cual excluye a las proposiciones alternativas A y J. En otras palabras, *una creencia es una proposición que es considerada y adoptada como verdadera por un sujeto*.

La ventaja que presenta esta definición de Price y Wilson radica justamente en que parte de la oposición entre creencia y conocimiento. En nuestro ejemplo, la creencia de Bety en C supone que Bety se inclina o asiente que C, esto es, Bety la considera verdadera, por lo cual rechaza las otras proposiciones. Pero C podría ser

falsa. De este modo, la condición de toda creencia es que se halla siempre la posibilidad de que la proposición sea falsa, aunque el sujeto la adopte como verdadera. Por el contrario, hablar de conocimiento supone que la proposición sea considerada, en primer término, verdadera. Aquí es donde justamente se presentan las controversias epistemológicas, debido a que los criterios de verdad y justificación que se han propuesto están abiertos a discusión entre los filósofos.

Por ello, se proponen los siguientes criterios para caracterizar, de forma provisional, al conocimiento: *una creencia verdadera se considerará conocimiento si una comunidad de expertos disponen de un conjunto de razones o de cierta evidencia para validar esta proposición, con lo cual se descartan otras proposiciones alternativas*. Evidentemente, esto no resuelve el problema de la justificación epistémica, pero puede servir como punto de partida para el análisis ulterior.

Pasemos ahora a definir qué es una creencia *folk*. Nuestro procedimiento sigue la propuesta de Price y Wilson: afirmamos que S tiene una creencia *folk* si S adopta una proposición p y p se encuentra en oposición a una proposición q , las cuales se refieren a un mismo hecho del mundo, pero que, a diferencia de p , q está respaldada por razones suficientes para ser considerada conocimiento por los expertos en el área en cuestión. Esto lo podemos expresar del siguiente modo:

La creencia en p es una creencia *folk* si:

- S cree que p .
- Una comunidad de expertos cree que q .
- p se opone a q .

El siguiente ejemplo permitirá ilustrar este planteamiento. Pedro cree firmemente que la Tierra es plana (T), lo cual se halla en oposición a la proposición de que la Tierra es un esferoide (E) sostenida por la astronomía. Entonces, la creencia de Pedro que afirma que la Tierra es plana es una creencia *folk* si:

- Pedro cree que T.

- La comunidad de astrónomos cree que E.
- T se opone a E.

Tanto T como E describen el mismo hecho del mundo, en este caso, la forma de nuestro planeta. Pero E es considerada conocimiento en la medida en que los astrónomos han mostrado cierta evidencia científica a su favor: los experimentos de comparación de sombras en diferentes latitudes realizados por Eratóstenes, el aparente hundimiento de los barcos cuando se alejan en alta mar, la sombra de la Tierra proyectada en la Luna durante un eclipse lunar, las fotografías tomadas por satélites artificiales, etc. De esta manera, la oposición entre T y E radica en que E es aceptada por los astrónomos como conocimiento, en tanto que T no lo es.

No obstante, las distinciones hasta ahora establecidas son insuficientes, ya que no aportan un criterio para distinguir entre las creencias posiblemente falsas sostenidas por un individuo y las creencias *folk*: si Pedro cree que el alma de su abuelo transmigró a su perro, podrá ser una creencia falsa, mas no sería necesariamente una creencia *folk*. Pero si una comunidad de individuos adopta la creencia de que las almas de las personas transmigran a las mascotas, esto sí conformaría una creencia *folk*. En este sentido, las creencias *folk* presentan un importante rasgo distintivo: su carácter social.

Las creencias *folk* son compartidas por una comunidad dada, que podemos denominar “público lego”. El público lego es aquel que carece de formación en una determinada área o disciplina de conocimiento. En nuestro ejemplo, T sería una creencia *folk* en la medida en que T es asentida o adoptada como verdadera por cierta comunidad que carece de formación en astronomía.

La definición de creencia *folk*, entonces, sería la siguiente: *una proposición, asentida o adoptada por el público lego, que se encuentra en oposición con otras proposiciones, asentidas o adoptadas como conocimiento por los expertos en un área o disciplina de conocimiento dada.*

¹ Las ciencias *folk* no deben ser confundidas con las llamadas pseudociencias (como la astrología, la parapsicología, el estudio del fenómeno de los ovnis, la homeopatía, etc.). Las primeras compiten con las disciplinas científicas establecidas en tanto que ofrecen explicaciones intuitivas de ciertos

De este modo, las llamadas “ciencias *folk*”¹ estarían conformadas por: 1) un conjunto de creencias que describen ciertos hechos que son explicados alternativamente por ciertas teorías científicas; 2) estas creencias contradicen las explicaciones planteadas por dichas teorías, y 3) existe una comunidad de legos que han adoptado tales conjuntos de creencias. Visto así, denominar ‘teoría’ a la postura terraplanista es una concesión excesiva, pues no se trata realmente de una teoría científica sino, en el mejor de los casos, de una creencia propia de la “física *folk*”. Lo mismo podría decirse del creacionismo y otras tantas creencias a las que se les concede injustificadamente el estatus de ‘teorías’.

Una vez esclarecidos los conceptos básicos de nuestro análisis, pasaremos ahora a revisar cuáles son las creencias económicas *folk* más frecuentes.

3. LA ECONOMÍA *FOLK* FRENTE A LA CIENCIA ECONÓMICA

Desde el surgimiento de la economía moderna, se han dado varios intentos por divulgar la ciencia económica al público lego (Bastiat 2004; Newcomb 1893; Hazlitt 1946; Friedman & Friedman 1980; Sowell 2015; Schettino 2002, 2015). No obstante, pese a estos diversos esfuerzos, la divulgación de la economía ha tenido muy poco éxito. Esto no es de extrañar, pues algo similar ha ocurrido con los programas de divulgación en física, biología, lingüística y otras disciplinas científicas.

Esto ha llevado a algunos teóricos a tratar de indagar los motivos del fracaso de la divulgación económica, llegando así a toparse con la brecha que se da entre

fenómenos naturales o sociales de los que se disponen ya de ciertas explicaciones científicas. En contraste, las pseudociencias pretenden imitar a las teorías científicas en cuanto a sus métodos y procedimientos, pero sus objetos de estudio son, en principio, dudosos: la influencia de los astros en el destino de las personas, las supuestas visitas de seres extraterrestres, la existencia de fantasmas y espíritus, etc. Así, pese a que las pseudociencias intentan dar ropaje de cientificidad a sus planteamientos, su problema fundamental se halla incluso a nivel descriptivo. Desde luego, muchas creencias *folk* pueden aludir a supuestos fenómenos sobrenaturales, por lo cual la línea divisoria entre ciencias *folk* y pseudociencias parece ser difusa. Esto requerirá investigarse más a futuro.

las concepciones de los legos y el conocimiento científico: en este caso, las teorías económicas. Bryan Caplan (2001, 2002a, 2002b, 2006, 2007), por ejemplo, ha señalado que las personas sin formación económica presentan ciertos sesgos cognitivos que se manifiestan cuando intentan explicar el funcionamiento del sistema económico, si bien este autor no emplea el término *folk belief*. Pero investigaciones más recientes (Rubin; Boyer y Petersen) son las que han propuesto el término ‘creencias económicas *folk*’ (*folk-economic beliefs*) y muestran resultados que se hallan en la misma sintonía de los estudios de Caplan antes referidos. De esta forma, existe ya una cierta cantidad de investigaciones acerca de las creencias económicas *folk*. La cuestión ahora es determinar en qué consisten tales creencias económicas, las cuales conforman lo que podríamos designar como ‘economía *folk*’.

Evidentemente, abordar esta cuestión no consistiría simplemente en enlistar las creencias que posee el público lego en materia de economía. Por tal razón, conviene precisar los rasgos definitorios de estas creencias a partir de la oposición que presentan con las proposiciones admitidas como conocimiento en la ciencia económica.

En principio, es necesario reconocer que en la ciencia económica existen múltiples debates respecto a ciertos problemas en particular (por ejemplo, en temas como las propuestas para enfrentar las externalidades o aspectos muy concretos de las políticas monetarias). No obstante, hay también ciertos principios básicos admitidos por consenso por la mayoría de los investigadores que conforman el *mainstream* de la ciencia económica (Ayau 2003; Parkin 2015; Gwartney et ál. 2016; Sowell 2015; Mankiw 2021). Podemos destacar los siguientes:

- A1) La riqueza de los países no se deriva de la cantidad de dinero o metales preciosos que posea un país, sino de la productividad de sus habitantes.
- B1) La satisfacción del interés personal contribuye al bienestar general. El mercado representa un juego de suma positiva (dado un intercambio económico, ambas partes se benefician mutuamente; es una situación de ganar-ganar).

- C1) Tanto para los individuos como para las naciones, especializarse en producir aquellas cosas para las que se tienen mejores capacidades o recursos es lo óptimo, y comprar las demás cosas a otros individuos o países, pese a que estas cosas también podrían ser producidas domésticamente (*principio de la ventaja comparativa*).
- D1) El sistema de precios informa sobre la relación oferta/demanda y permite la coordinación de las actividades económicas en el mercado, sin necesidad de una autoridad central planificadora.
- E1) El valor de bienes y servicios es subjetivo. No es una propiedad objetiva de las cosas, sino que depende exclusivamente de los deseos y las preferencias de los individuos.
- F1) El valor de bienes y servicios es marginal y relativo. No es una propiedad absoluta, sino que depende de la posesión, el uso o el consumo previos de esos mismos bienes.

Estas son algunas de las proposiciones básicas aceptadas como conocimiento por consenso entre la mayoría de economistas. La economía clásica estableció los principios que van de A1 a D1, en tanto que la economía neoclásica incluyó E1 y F1. A partir de estas proposiciones, podemos extraer algunas implicaciones teóricas. En primera instancia, la ciencia económica parte de la primacía del individuo al colocarlo como el factor central de los procesos económicos: los individuos buscan satisfacer sus deseos y preferencias subjetivas, lo cual genera incentivos para que consumidores, intermediarios y productores interactúen entre sí. Dada la escasez relativa de recursos que poseen usos alternativos, es preciso que estos sean asignados de la forma más eficiente posible. Por ejemplo, el algodón puede ser empleado para producir ropa, cuerdas, material clínico, biocombustibles, etc. Cuántas cantidades de algodón se enviará a cada una de las industrias que producen estos materiales y de qué forma se coordinarán los agentes económicos para efectuar estas actividades es el gran reto de todo sistema económico, y es también el objeto de estudio central de la ciencia económica.

El objetivo de los intercambios económicos es, entonces, conseguir la asignación más eficiente de recursos escasos para producir bienes y servicios, la cual se logra mediante el sistema de precios en las economías de mercado. Este actúa como un medio de información acerca de la demanda de los consumidores y la disponibilidad de recursos escasos (oferta) para satisfacer esa demanda. En este sentido, los precios se determinan a partir de cuántas unidades adicionales (“marginales”) de los bienes puedan ser generadas por los productores y puedan ser adquiridas por los consumidores. Del mismo modo, los precios generan incentivos que estimulan a los oferentes a producir ciertos bienes y que obligan a los consumidores a modificar sus gastos. Así es cómo, en términos muy generales, se afronta el problema de la asignación de recursos escasos con usos alternativos y la coordinación de las actividades económicas.

En contraste con estos principios, algunas de las creencias económicas *folk* más extendidas son las siguientes:

- A2) La riqueza de los países se deriva de la cantidad de dinero o metales preciosos poseídos.
- B2) La satisfacción del interés personal solo contribuye al enriquecimiento personal. El mercado representa un juego de suma cero (dado un intercambio económico, una de las partes se beneficia a expensas de la otra; es una situación de ganar-perder).
- C2) Lo mejor para un país es reducir sus importaciones e incrementar sus exportaciones. La economía nacional debe aspirar a la autonomía (*principio de autarquía*).
- D2) El sistema de precios refleja la avaricia de los intermediarios y es necesaria la intervención gubernamental mediante el control de precios.
- E2) El valor de bienes y servicios es una propiedad objetiva intrínseca, y es producida por la cantidad de trabajo que se invierte en ellos.
- F2) El valor de bienes y servicios es absoluto: no depende de las circunstancias o los valores de otros bienes.

Como se puede apreciar, estas creencias *folk* se hallan en franca oposición con los principios económicos mencionados. Específicamente, las oposiciones se presentan de la siguiente manera:

- La oposición entre A1 y A2 radica, en primera instancia, en una diferencia semántica. La definición del término “riqueza” en la ciencia económica es distinta a la del uso coloquial del término, como se puede apreciar en la entrada del diccionario en línea de la Real Academia de la Lengua Española:
 1. f. Abundancia de bienes y cosas preciosas.
 2. f. Abundancia de cualidades o atributos excelentes.
 3. f. Abundancia relativa de cualquier cosa. *Riqueza alcohólica, de minerales, de vocabulario.*

De las tres acepciones, solo la primera parece relacionarse con la concepción económica. Es preciso hacer notar que se define ‘riqueza’ como la posesión de bienes y ‘cosas preciosas’, lo que muestra claramente que esta definición corresponde con la creencia A2. En economía, desde Adam Smith (en su clásico *Wealth of Nations* 1976), se establece que la riqueza se refiere a los bienes y servicios que son producidos para la satisfacción de los deseos individuales. Si bien es cierto que la cantidad de dinero que posea un individuo le permitirá obtener mayores cantidades de bienes que satisfagan sus deseos, el dinero en sí no conforma la riqueza, pues simplemente se trata de un medio de intercambio para obtener dichos bienes. De hecho, si la riqueza de un país estuviese definida solo por la cantidad de dinero, bastaría con imprimir más circulante para hacer que el país en cuestión se volviese rico. Pero, como han advertido en reiteradas veces los economistas, el resultado de imprimir grandes cantidades de dinero solo provoca presiones inflacionarias que afectan negativamente a las economías nacionales.

De esta forma, no es la cantidad de dinero o ‘cosas preciosas’ lo que caracteriza a la riqueza, sino la producción de los bienes y servicios. Por ello, como ya destacaba Smith, lo que determina el incremento de la riqueza es la productividad.

En contraste, la creencia A2 asume que la riqueza es algo que ya está dado: se concibe a la riqueza como un conjunto finito de bienes (metales preciosos, productos agrícolas, petróleo, etc.) que ha existido desde el inicio de los tiempos y, desde entonces, los individuos o grupos sociales han luchado entre sí para obtener una parte de ella (Pinker 2018 80). Como se verá, las siguientes creencias *folk* guardan una estrecha relación con esta, pues así como el principio A1 es la base de la teoría económica moderna, A2 representa lo que podríamos denominar la “creencia económica madre”.

- La oposición entre B1 y B2 está estrechamente ligada a lo expuesto. La idea de que el mercado es un juego de suma cero (Rubin 2003 157ss; Sowell 2011 3ss) se puede ilustrar con la metáfora del pastel propuesta por Milton Friedman. Si la riqueza se puede imaginar como un pastel con un tamaño fijo —esto es, que la riqueza representa siempre una cantidad fija (A2)—, se concibe que si un individuo o grupo de individuos recibe una tajada mayor del pastel, el resto de la población recibirá proporcionalmente tajadas menores. De ahí que se infiera que la principal causa de los problemas sociales se derive de un “reparto injusto” del pastel.

La cuestión es que el mercado no es algo fijo ni la economía es un juego de suma cero: no es un pastel, sino una maquinaria que produce masivamente pasteles. Si los intercambios económicos consistiesen en juegos de suma cero, las partes involucradas simplemente no participarían en los intercambios. Por otra parte, si en una economía de mercado el incremento de la riqueza se presentase a expensas de grandes sectores de la población, esto se mostraría empíricamente a partir de una correlación entre el crecimiento económico y el aumento de la pobreza. Sin embargo, el espectacular crecimiento económico y la drástica reducción

de la pobreza en China —a partir de 1978— y en la India —a partir de 1991— muestran una correlación en el sentido contrario (Bhagwati & Panagariya 2013; Ang 2016; Bhalla 2018). Estos casos muestran que la elevación de los estándares de vida de grandes sectores de la población y la apertura económica van aparejados (Deaton 2013; Lal 2013). Por lo tanto, la creencia de que el mercado es un juego de suma cero es simplemente falsa.

Por otro lado, la creencia B2 parte de la presunción de que únicamente los actos desinteresados, como los actos altruistas, pueden ser socialmente benéficos. De esto se sigue que las acciones que pretendan satisfacer el interés personal solo pueden generar beneficios a las personas que las realizan. Lo que está detrás de esta creencia es una cierta valoración moral, pues en esta creencia está implícita una censura de la búsqueda de la satisfacción personal: si una persona se enriquece, entonces es avara y egoísta. En una forma más sofisticada, algunos teóricos han acusado a la economía de promover el egoísmo (Heyne 2008 1ss; Coleman 2002 136ss). En suma, el problema en B2 es que excluye la posibilidad de que la búsqueda de la satisfacción personal que, según los economistas, es la motivación principal para el intercambio económico, pueda ser socialmente benéfica (Ayau 2003 2ss).

- La oposición entre C1 y C2 presenta una peculiaridad. Los individuos suelen aplicar la ventaja comparativa (Gwartney et ál. 2016: 36) en lo que respecta a sus actividades personales (por ejemplo, aunque una persona sea muy diestra para hacer el aseo de su hogar, podrá optar por contratar a otra persona que realice tal actividad, puesto que esto le permitirá ahorrar tiempo que puede dedicar a las actividades que le resultan más importantes o en las que es más productivo) y pueden no tener problemas en aceptar teóricamente este principio en lo tocante a los asuntos personales, pero lo rechazan cuando se refiere a cuestiones públicas, como la economía nacional. Es decir, los legos pueden aceptar

C1 cuando se trata de sus intereses privados, mas aceptan C2 cuando se trata del ámbito público. Ejemplo de ello es la consideración de que las importaciones afectan a la economía nacional o la creencia de que las economías nacionales deberían aspirar a la autarquía. Pero, como señalan los economistas (Mankiw 2021 52), los costos de abarcar muchas actividades económicas variadas exceden a los posibles beneficios, ya que los recursos se estarán destinando a producir bienes que resultan más costosos de producir localmente. Por ello, lo más benéfico para una economía sería enfocar sus recursos en la producción de aquellos bienes que implican menos costos o generan más utilidades, esto es, lo óptimo es aprovechar las ventajas comparativas de un país. Una muestra de que el público lego entiende la ventaja comparativa se puede apreciar en el viejo y conocido refrán “El que mucho abarca, poco aprieta”; la cuestión es que parecen no aplicarlo más allá de su esfera privada.

- En cuanto a la oposición entre D1 y D2, no solo radica en una perspectiva parcial y limitada del funcionamiento del sistema de precios implícita en D2, sino de otro supuesto: que la intervención gubernamental es el mejor recurso para beneficiar a la sociedad. En la perspectiva del lego, si los intercambios son juegos de suma cero y la satisfacción del interés personal supone acciones egoístas, se asume que el Estado es un tercero que permite evitar el abuso o la explotación por parte de ciertos agentes económicos, como los intermediarios (Sowell 2011; Wheelan 2019). El problema es que el lego está asumiendo, con ello, que este tercero posee toda la información necesaria para conseguir la asignación eficiente de bienes y servicios, y, al mismo tiempo, representa la autoridad moral para presuntamente lograr el bienestar general —en otras palabras, se presupone que el Estado es omnisciente y altruista—. Mas el control de precios y otras medidas intervencionistas, según los economistas, distorsionan la información que se genera en los intercambios económicos por medio del sistema de precios (Sowell 2015 92; Gwartney et ál. 2016 186). Por otra parte, los individuos que laboran en las

instituciones gubernamentales también buscan satisfacer sus intereses personales, como cualquier agente económico (Buchanan). De ahí que resulte difícil sostener el supuesto altruísmo de los gobernantes.

- En lo que respecta a la concepción subjetivista del valor (E1), resulta claramente incompatible con la concepción objetivista (E2). Esta última no solo está arraigada en el público lego, sino que los economistas clásicos la habían aceptado (Adam Smith, David Ricardo, François Quesnay, Thomas Malthus, John Stuart Mill, etc.) al no disponer de una teoría alternativa en su época. Esto se debió, en gran medida, a que los economistas clásicos estudiaron los fenómenos económicos en términos de “clases” de bienes y valores, y no de preferencias individuales (Rothbard 1983 20-27). No sería sino hasta la revolución marginalista, iniciada por William S. Jevons (1862) y Leon Walras (1874), y desarrollada ampliamente por los austriacos Carl Menger (1871) y Eugen von Böhm-Bawerk (1884), cuando se desarrolló plenamente la concepción subjetivista del valor y el análisis marginal, en adelante reconocidos como los pilares teóricos de la economía neoclásica.

En todo caso, el público lego asume que hay “valuaciones objetivas” de los bienes o servicios, esto es, que los objetos poseen un valor inherente, independiente de las preferencias y deseos de los individuos. Pero, como señaló Menger, cada individuo posee una “lista de preferencias” con la cual los consumidores clasifican subjetivamente los bienes según el grado de satisfacción (*utilidad*) que obtengan de ellos. Si se contrastan las “listas de preferencias” de diferentes individuos, se puede constatar que varían notablemente respecto a qué clase de bienes son incluidos en las listas de unos y otros; incluso, aunque algunos bienes coincidan en el listado de dos individuos, podrán aparecer en diferentes posiciones en sus respectivas listas, según el grado de utilidad que estos esperan de ellos.

- Sin embargo, las valuaciones no solo son subjetivas y difieren de un individuo a otro (E1), sino que pueden cambiar en un mismo individuo bajo diversas circunstancias (F1): un consumidor podrá valorar más un helado de limón en Cancún que en las zonas gélidas de Canadá; podrá valorar más un café por la mañana que por la noche; podrá valorar más el ahorro en su etapa adulta que en su juventud, etc. En consecuencia, el valor de un bien o servicio dependerá de cuántas unidades marginales estén disponibles y qué cantidad de estas el consumidor está dispuesto a adquirir, dadas ciertas circunstancias. En contraste, el público lego cree que los bienes tienen un valor absoluto (F2): el café o el helado de limón son valiosos siempre y en todo momento.

Lo decisivo en este punto es que el análisis marginal permite explicar múltiples fenómenos económicos. Por ejemplo, algunos observadores externos podrán pensar que un restaurante mejoraría mucho su servicio si el dueño contratara más personal. Pero el dueño debe considerar si la contratación de más trabajadores contribuye a la productividad de su negocio: si la utilidad marginal de un nuevo empleado supera su costo marginal, entonces valdrá la pena pagar un salario extra; si no es el caso, lo más conveniente para su negocio es no aumentar la planta laboral. Después de todo, el número de clientes y la carga de trabajo será mayor o menor en ciertos horarios y en ciertos días de la semana, factores que considera el empleador antes de tomar este tipo de decisiones. En suma, entender las situaciones particulares en términos marginales permite explicar con mayor precisión los fenómenos en vez de adoptar un enfoque absoluto y rígido, propio de las concepciones legas sobre la economía.

De esta manera, se presentaría el conjunto de oposiciones entre la ciencia económica y las creencias que integran la economía *folk*. Cabe

destacar que algunas de estas creencias históricamente precedieron al surgimiento de la economía como disciplina científica, como las creencias esencialistas y vitalistas precedieron a la actual biología evolutiva, o las concepciones geocéntricas antecediéron a la astronomía moderna. Por ejemplo, las creencias A2 y B2 fueron sostenidas por los mercantilistas (siglos XVI al XVIII), cuyas ideas influyeron decisivamente en las políticas en contra del libre comercio en aquellos tiempos. Los mercantilistas planteaban que la riqueza de una nación consistía en su inventario de metales preciosos, por lo que recomendaban que un país redujera sus importaciones e incrementara sus exportaciones, ya que la compra de más productos importados implicaría una reducción de los activos en las arcas nacionales. En todo caso, gran parte del esfuerzo teórico de los economistas clásicos consistió justamente en combatir esta doctrina. Paradójicamente, las recetas mercantilistas fueron resucitadas por la Cepal en la década de los cincuenta del siglo pasado, lo cual evidencia la persistencia de ciertas creencias equivocadas. Esta cuestión será atendida líneas más adelante.

4. POSIBLES EXPLICACIONES DE LAS CREENCIAS ECONÓMICAS *FOLK*

A partir de lo expuesto previamente, surge ahora la cuestión de por qué los legos adoptan tales creencias. Como posibles respuestas a esta pregunta, presentaré a continuación tanto las siguientes hipótesis, como los respectivos argumentos que pretenden mostrar por qué considero que tales hipótesis podrían explicar este fenómeno:

H1) *Las creencias económicas folk son causadas por las apreciaciones limitadas de los fenómenos por parte del público lego.* Las personas tienden a percibir e interpretar los fenómenos a partir de sus propias circunstancias. Como ocurre con otras ciencias, los fenómenos que describe e intenta explicar la ciencia económica suponen cierto grado de abstracción, dado que son complejos procesos sistémicos

que trascienden el ámbito local en el que se desenvuelven los individuos y resultan complicados de comprender sin las herramientas teóricas de la ciencia económica. Bienes que resultan tan comunes y necesarios para los individuos como la ropa, los electrodomésticos, los teléfonos móviles o celulares y los alimentos son resultado de la coordinación de miles o millones de agentes económicos que no se conocen entre sí y probablemente nunca lleguen a verse las caras. Tal coordinación no ha sido planificada ni ordenada por ningún señor feudal o ningún comité central, sino que emerge a través de esa infinidad de interacciones individuales. Por ello, muchos de los principios económicos fundamentales resultan contraintuitivos para los legos y, entonces, desafían sus concepciones acerca de cómo funciona la sociedad.

Quizás esto permita explicar la popularidad de las mal llamadas “teorías conspirativas” —otro ejemplo del uso indebido del vocablo “teoría”—, que afirman que la economía mundial es controlada por cofradías ultrasecretas. Desde la perspectiva del lego, resulta impensable que los hechos económicos dependan de procesos sistémicos que emergen de los innumerables intercambios particulares, lo que implica la suposición de que debe existir alguna clase de control oculto que coordina la ingente actividad económica para algún propósito perverso. Lo que subyace a esta forma de pensar es la idea de que todos los eventos son resultado de actos intencionales, y esto es un rasgo común de muchas concepciones ingenuas, no solo relativas a la economía: en tiempos pasados, se solía pensar que fenómenos naturales como la lluvia, los terremotos o los eclipses eran deliberadamente causados por los dioses u otros entes sobrenaturales. El problema de las “teorías conspirativas” radica en que parten de supuestos falsos: 1) que los conspiradores secretos poseen todo el conocimiento que circula en la economía (lo cual es imposible, dados los múltiples factores que inciden en los hechos económicos), y 2) que en el mundo no existe la casualidad. Lo curioso es que la ciencia económica, a lo largo de sus casi doscientos años de existencia, ha venido desarrollando paulatinamente ciertos modelos teóricos y herramientas metodológicas para intentar explicar, con la mayor precisión posible, el funcionamiento de la economía, mientras que los “teóricos” de la conspiración disponen ya de todas las respuestas. Como sea, estas seudoteorías podrían tener su origen en ciertas concepciones de la economía *folk*.

Ya el economista clásico Frédéric Bastiat (2004) había mostrado que muchas confusiones en materia económica se derivan de una visión limitada de los individuos, que tienden a considerar únicamente los beneficios de ciertas decisiones —“lo que se ve” — pasando por alto sus costos, o bien, ignorando las consecuencias no intencionadas de ciertas acciones —“lo que no se ve” —. Por ejemplo, Bastiat cuestiona la creencia de que los impuestos son en sí mismos benéficos para la economía, argumentando que si bien ciertos sectores se pueden ver favorecidos por el gasto gubernamental, los costos invisibles que impone la tributación pueden exceder tales beneficios, lo que puede afectar a la economía en su totalidad. Así pues, las creencias *folk* podrían derivarse de una apreciación limitada y parcial del funcionamiento de la economía.

H2) *Las creencias económicas folk son causadas por valoraciones morales.* Es usual que las personas emitan juicios de valor antes de entender ciertos fenómenos. Esta tendencia a confundir explicaciones con valoraciones ya había sido observada por el filósofo René Descartes (2011 *Cuarta Meditación*), quien argumentaba que una causa frecuente del error se deriva de que la voluntad se anticipa al entendimiento: aceptamos algo como verdadero porque nos resulta agradable (o moralmente deseable) y, *mutatis mutandis*, rechazamos algo como falso en tanto que nos resulta desagradable (o moralmente indeseable). La cuestión es que la verdad o falsedad de las proposiciones es indiferente a las valoraciones morales.² Por ejemplo, el aumento de precios suele dar lugar a quejas de los consumidores, quienes acusan a los intermediarios de ser abusivos y de querer “lucrarse” con las necesidades de la gente.

Supongamos que una persona desea alquilar un departamento en la ciudad X, pero encuentra que los precios son muy altos en ciertas zonas de la ciudad. Podrá recriminar a las empresas de bienes y raíces de ser “avaras” y de cobrar precios “excesivos”. Pero si en algunas zonas de la ciudad los precios son menores, ¿significa entonces que la avaricia “es menor” en esas partes de la ciudad? El hecho de que los precios

² Sobre las diferencias entre moral y ciencia, y, más concretamente, entre moral y economía, véase Coleman (2002 119-131).

de los inmuebles (para venta o alquiler) sean mayores en el barrio 1 que en el barrio 2 indica que, por una parte, tales bienes son relativamente escasos y, por otra parte, que hay mayor demanda de casas y departamentos en 1 que en 2. Así, no es difícil entender que los precios de inmuebles sean comparativamente más bajos en barrios peligrosos (como podría ser el caso del barrio 2) que en otras partes de la ciudad. La falta de una comprensión mínima de la dinámica de los precios —concretamente, de los factores que inciden en la oferta y la demanda— puede llevar a formular explicaciones erróneas, cargadas de valoraciones morales precipitadas.

Anteriormente, se señaló que muchas creencias dan lugar a ciertos juicios moralistas, aunque la relación causal podría ser a la inversa: las creencias de los legos son el resultado de esta confusión entre explicación y valoración moral. Si este es el caso, es posible que las creencias económicas *folk* se deriven de ciertas escalas de valores previamente asumidas por los individuos, valores que, a su vez, podrían ser inculcados por la tradición o por la formación religiosa —por ejemplo, la célebre sentencia bíblica: “Yo os aseguro que un rico difícilmente entrará en el Reino de los Cielos [...] es más fácil que un camello entre por el ojo de una aguja, que el que un rico entre en el Reino de los Cielos”—(Mateo 19 23-30). Por ende, es posible que ciertas valoraciones morales den lugar a ciertas concepciones económicas legas.

H3) *Las creencias económicas son causadas por la especialización cognitiva de los individuos en determinadas actividades o áreas del conocimiento.* Esta hipótesis se basa en dos hechos fundamentales: 1) los individuos poseen capacidades y habilidades distintas (sean congénitas o adquiridas), y 2) los individuos disponen de un tiempo y energía limitados, por lo cual deciden invertir su tiempo y energía solamente en desarrollar esas capacidades y habilidades (Ayau 2003 2-ss). Supongamos el caso de un cirujano, quien ha logrado convertirse en un médico prestigioso en su especialidad. Él ha dedicado mucho tiempo y esfuerzo en adquirir los conocimientos de su disciplina, la ha practicado durante años, desarrollando con ello ciertas destrezas, y continuamente debe informarse sobre los avances más recientes en su área de especialidad; se trataría, pues, de una “ventaja comparativa epistémica”. Esto le deja poco tiempo para informarse en otras áreas de conocimiento u otras actividades, como la teoría política o económica. Para el cirujano resultaría muy costoso dedicar tiempo

al estudio de estas u otras áreas, ajenas (directamente) a su actividad, pues implicaría una inversión de tiempo y esfuerzo que bien podría dedicar a su propia disciplina. Por ello, posiblemente el cirujano podría adoptar creencias económicas *folk*, como creer que lo mejor para el país es adoptar un modelo basado en la autarquía. De igual forma, un economista podría formar creencias médicas *folk* —v. gr., creer en remedios dudosos como que el dióxido de cloro “cura” el Covid-19— básicamente por la misma causa: la especialización cognitiva en su área de conocimiento. En suma, no podemos conocerlo todo dadas nuestras limitaciones humanas y, en consecuencia, nos enfocamos en aprender los conocimientos que son relevantes para nuestras actividades (esto es, la “ventaja comparativa epistémica”). Tal vez esta sea una causa de la adopción de este tipo de creencias económicas.

H4) *Las creencias económicas son causadas por sesgos cognitivos.* Recientes estudios en psicología cognitiva (Kahneman et ál. 1982; Kahneman et ál 2002; Baron 2007; Evans & Frankish 2012) han abierto un vasto campo de investigación, que se ha extendido a la medicina, la política y la economía, acerca de errores sistemáticos que cometen los individuos en sus procesos de razonamiento que pueden estar vinculados a la toma de decisiones. Según estas investigaciones, el ser humano procede cognitivamente de dos formas: 1) un proceso intuitivo, que consiste en reacciones inmediatas con las cuales se forman juicios ante ciertas circunstancias, y 2) un proceso racional, complejo y abstracto, que supone un mayor esfuerzo cognitivo con el cual se reflexiona acerca de conceptos y situaciones más generales. De este modo, estos “atajos mentales”, llamados sesgos cognitivos, resultan de aplicar intuiciones a ciertos objetos abstractos y complejos. Un caso sería la “disponibilidad heurística” (Pinker 2018 41-42), que consiste en emitir un juicio a partir de lo primero que se presenta en la memoria del sujeto. Este caso se puede apreciar en ciertas personas que consideran que es más seguro viajar por carretera que tomar un avión simplemente porque en las noticias se informó de un accidente aéreo reciente. No obstante, las estadísticas indican que son más frecuentes los accidentes por vía terrestre y que, de hecho, resulta más seguro trasladarse por vía aérea.

Los estudios de Caplan (2002b) sugieren que las concepciones de los legos en economía podrían ser explicadas en términos de sesgos cognitivos. Un ejemplo po-

dría ser el “sesgo pesimista” (*pessimistic bias*), según el cual los sujetos conciben que el desempeño de la economía ha empeorado año tras año: los legos pueden llegar a esta conclusión a partir de que una compañía importante se fue a la bancarrota, de que los precios han aumentado o de que personas cercanas han perdido su empleo, pero las estadísticas pueden indicar lo contrario. Por supuesto, estos individuos pueden desestimar las estadísticas, lo cual puede ser resultado de que consideran más fiables sus propias intuiciones. De esta manera, los sesgos cognitivos podrían aportar un modelo causal que explicaría la adopción de creencias económicas *folk*.

H5) *Las creencias económicas son causadas por una asimetría entre decisiones privadas y decisiones públicas.* El sociólogo noruego Jon Elster (1991 40) explora el planteamiento marxista de que, a lo largo de la historia, han existido contradicciones sociales, señalando que los individuos pueden poseer creencias y deseos contradictorios. Por ejemplo, un capitalista puede desear que sus obreros obtengan bajos salarios porque esto lo beneficia, y, al mismo tiempo, puede desear que los otros capitalistas paguen altos salarios a sus obreros ya que así estos podrán comprar sus bienes o servicios. Otro ejemplo: algunos empresarios pueden enarbolar una retórica a favor del libre mercado pero negarla en la práctica cuando cabildean en las esferas políticas para lograr la aprobación de leyes o políticas en beneficio de sus propios intereses, como exigir aranceles que protejan sus negocios frente a la competencia extranjera o solicitar subsidios gubernamentales (Hutt 1971; Buchanan 1999).

No obstante, esta asimetría entre decisiones privadas y públicas no se halla exclusivamente en las élites intelectuales o en las poderosas cúpulas empresariales, porque en todos los agentes económicos puede existir un conflicto entre lo que consideran mejor para sí mismos y lo que consideran mejor para los demás. En el apartado anterior, señalé que la ventaja comparativa suele ser aceptada por las personas cuando se trata de sus actividades personales, pero puede ser rechazada cuando se refiere a la *Res Pública*. Otro caso de esta asimetría podría ser el de los juicios respecto a los impuestos al ingreso. Muchas personas sostienen que debería implementarse una tributación progresiva para “distribuir la riqueza”, de tal modo que los más ricos paguen una cuota más alta, pero estas mismas personas suelen quejarse amargamente cuando su ingreso salarial se reduce considerablemente por las retenciones fiscales

como el impuesto sobre la renta (ISR). Estos ejemplos indican que podría existir un cierto patrón que induce a los individuos a caer en ciertos comportamientos contradictorios, que los llevan a concebir los fenómenos económicos de cierta forma cuando se refiere a su esfera privada y concebirlos de una forma distinta respecto de la esfera social. Así, la asimetría entre las decisiones públicas y privadas podría ser la causa de estas creencias.

Las hipótesis anteriores podrían ser complementarias. H1 y H3 se complementan en tanto que la observación parcial de los fenómenos económicos apunta hacia el conocimiento escaso o nulo, que puede ser resultado de la “ventaja comparativa epistémica”. Ambas también son compatibles con H2, pues la observación parcial puede estar entremezclada con valoraciones morales, como sugiere un poco el ejemplo acerca de la avaricia de los intermediarios. De igual manera, es posible que el lego en economía evite mezclar explicaciones con valoraciones morales en su campo de especialidad, pero no lo evite en campos que le resultan ajenos: por ejemplo, que nuestro cirujano evalúe cierto tratamiento por su eficacia y no por su posible valor moral.

En cuanto a H5, también podría complementar las hipótesis anteriores: las valoraciones morales pueden estar vinculadas a la manera como los sujetos perciben y evalúan el ámbito público,³ mas no las aplican a sus propias decisiones, lo que puede derivar en contradicciones performativas (por ejemplo, una persona puede argumentar que lo mejor es comprar a las productores nacionales porque así se les beneficia frente a la competencia extranjera, pero, al mismo tiempo, esta misma persona puede ser dada a comprar productos importados). De esta forma, la percepción parcial, la confusión entre valoraciones morales y explicaciones, la especialización en algunas áreas de conocimiento y la asimetría entre lo privado y lo público podrían estar estrechamente relacionadas.

Las hipótesis antes referidas también se complementan con H4 en la medida en que los sesgos cognitivos podrían apuntar hacia ciertos mecanismos que se dan en la cognición, que podrían ser la causa última de este tipo de creencias y las consecuentes decisiones (inconsistentes, en algunos casos). Evidentemente, si las hipótesis realmente se complementan o se contradicen entre sí, esto será determinado a partir de una investigación empírica.

No obstante, existe un planteamiento teórico que tal vez aglutine las hipótesis antes expuestas desarrollado por la psicología evolutiva (Pinker 2002; Rubin 2003; Boyer & Peterson 2018). Básicamente, esta teoría sostiene que posiblemente existe un desfase entre las capacidades del cerebro humano y las condiciones específicas de las sociedades actuales. La teoría parte de que nuestros cerebros evolucionaron en un entorno adaptativo muy diferente al que vivimos en el presente.

El argumento principal de esta teoría parte de la dimensión temporal. Nuestra especie tiene aproximadamente entre 350.000 y 300.000 años habitando este planeta. La civilización más antigua conocida data de hace 9000 años, y las actuales sociedades comerciales industrializadas tendrían su origen hace apenas unos 200 años. El cerebro humano no pudo haber evolucionado para responder a estas nuevas condiciones puesto que los cambios evolutivos requieren largos periodos de tiempo. En otras palabras, seguimos teniendo cerebros de cazadores-recolectores pese a que vivimos en sociedades más complejas y avanzadas respecto a las que habitaron nuestros antepasados. Pero, ¿en qué se caracterizaban las condiciones del hombre primitivo?

En aquellos tiempos, las interacciones sociales podrían haberse fundado en el altruismo recíproco, que se puede formular con el célebre dicho: “Hoy por ti, mañana por mí”. Dadas las condiciones extremas que enfrentaban los primeros *Homo Sapiens*, donde las fuentes de alimentación eran muy escasas, las interacciones económicas pudieron estar orientadas por esta clase de altruismo, que obligaba a todos los miembros de la comunidad a compartir lo cazado o lo recolectado, de tal modo que si algún miembro acumulaba más que los demás resultaba en una ganancia a costa del resto del grupo —en otras palabras, consistiría en un juego de suma cero—. Si este fuese el caso, las conductas egoístas podrían haber ameritado una sanción social: por ejemplo, que los miembros de la tribu se negaran a cooperar a futuro con los miembros que se comportaran de forma egoísta, lo que evidentemente pondría en peligro su propia supervivencia. La moral fundada en el altruismo recíproco es básicamente intencional: exige que los miembros de la comunidad conozcan las intenciones de sus camaradas para poder identificar si estos están dispuestos a cooperar o no. En consecuencia, el altruismo recíproco solo puede funcionar en pequeñas

sociedades, y no es de extrañar que persista en la actualidad en grupos sociales minúsculos, como las familias.

Sin embargo, las interacciones económicas en nuestra sociedad contemporánea son muy diferentes pues se basan en el comercio y la producción industrial, que generan condiciones propicias para juegos de suma positiva. Como destacamos, la coordinación de las actividades económicas en las grandes sociedades actuales supone las interacciones entre miles o millones individuos que se desconocen entre sí. En estas circunstancias, se han desarrollado sistemas jurídicos formales y marcos institucionales despersonalizados para poder regular el establecimiento de contratos y garantizar la seguridad jurídica de los agentes económicos. Es por ello que el altruismo recíproco de las tribus primitivas simplemente no tiene cabida a gran escala ya que es imposible rastrear las intenciones de miles o millones de agentes individuales que participan en estos vastos procesos sistémicos. Caracterizar a las sociedades modernas como “una gran familia” puede ser una metáfora retóricamente muy atractiva, pero es completamente inaplicable en términos descriptivos y explicativos.

En consecuencia, si este planteamiento teórico es correcto, es posible que la causa de las creencias económicas *folk* se derive de este desfase entre los modos de interacción de nuestros ancestros, que quedaron registrados en el cerebro humano, y la realidad socioeconómica actual. Como establecen los psicólogos evolutivos, hay dos tipos de explicación en términos evolutivos: las próximas (que responden a la pregunta: “¿cómo funciona X?”) y las últimas (cuya pregunta es: “¿por qué existe X?”). Las hipótesis H1, H2 y H3 se podrían entender como explicaciones próximas, ya que darían cuenta de ciertos factores específicos que causan el fenómeno, mientras que H4 se consideraría una explicación última, pues podría estar basado en un mecanismo general evolutivo que predispondría la formación de creencias económicas *folk*.

De igual forma, es posible que esta teoría permita comprender la conformación de ciertas doctrinas éticas asociadas con esta clase de creencias. Según Heyne (2008), nuestras intuiciones morales se caracterizan por lo que él denomina “moral de cara a cara” (*face-to-face morals*), la cual coincide con el altruismo recíproco, descrito líneas atrás. En este sistema moral, como ya se indicó, se pone el énfasis en las

intenciones de los agentes morales y se reprobaban las conductas egoístas. En este sentido, las doctrinas éticas propuestas por muchas religiones y por filósofos antiguos como Platón, Aristóteles o Séneca, en las que explícitamente se condena el “afán de lucro”, la “avaricia” y otras conductas, podrían tener un origen evolutivo. Quizá la diferencia que puedan tener estas doctrinas con las creencias *folk* podría radicar en que las primeras han sido formuladas de manera más sistemática y abstracta. De esto se sigue que el desfase entre nuestros cerebros de cazadores-recolectores y las condiciones propias de las economías de mercado modernas podría explicar la persistencia de la “moral de cara a cara” y las doctrinas filosóficas inspiradas en esta.

Sin embargo, esta teoría es todavía muy especulativa (Niemietz 2019 286). Como suele ocurrir con los planteamientos teóricos de la psicología evolutiva, el problema radica en las dificultades para hallar evidencia acerca de la filogenia de los procesos cognitivos y las condiciones de vida de nuestros ancestros primitivos. A pesar de ello, creo que esta teoría merece ser tomada en consideración, pues no está descartado que puedan encontrarse datos que la respalden en el futuro.

5. IMPLICACIONES POLÍTICAS

Al margen de si las creencias *folk* y las intuiciones morales de los legos en economía poseen raíces evolutivas, la cuestión ahora es ver qué repercusiones pueden tener para el ámbito político. Como señalé al inicio de este texto, las creencias no solo aportan descripciones del mundo a los sujetos con las cuales intentan afrontar la incertidumbre, sino que establecen un poderoso motor para la acción. Y las acciones que tienen efectos en la política y la sociedad en general cobran mayor relevancia ya que afectan las vidas de millones de personas. El impacto que las creencias y el conocimiento tiene para la política se ha convertido en uno de los temas de mayor interés para la epistemología y las ciencias cognitivas (Haidt 2012; Hannon & de Ridder 2021).

Sin embargo, no todas las creencias *folk* parecen tener el mismo impacto social. Las creencias lingüísticas *folk*, como las que describe Jackendoff, podrán llevar a los hablantes a formar visiones erróneas acerca de la lengua, pero no parecen tener

mayores repercusiones. En otros casos, los efectos negativos de algunas creencias parecen ser más claros: las concepciones legas en medicina pueden provocar que ciertos individuos tomen tratamientos inadecuados que pueden poner en riesgo su salud o su propia vida. Precisamente, el énfasis de los científicos en dar mayor divulgación al conocimiento científico y refutar públicamente las creencias equivocadas tiene que ver con la pretensión de informar a las personas para prevenir que sean víctimas de bulos o fraudes, que pueden afectar sus finanzas o incluso su salud. De este modo, el impacto negativo que puedan producir las creencias *folk* es variable, en tanto que parece depender de los asuntos específicos que están en juego (como la salud mental o física de las personas).

Pero el caso de las creencias económicas *folk* es distinto, porque atañen directamente a políticas públicas. Por ejemplo, la creencia de que la economía es un juego de suma cero ha producido ciertas posturas xenofóbicas, como la que afirma que los inmigrantes extranjeros “hurtan” empleos locales. Esto puede conducir al establecimiento de políticas discriminatorias y legislaciones draconianas que perjudicarían no solo a los inmigrantes, sino a la economía en su totalidad. Este y otros ejemplos son claros indicadores de que las concepciones económicas de los legos pueden imponer terribles costos a la sociedad.

Los economistas han sido conscientes de este problema, pero lo han desestimado, de cierta forma, apelando a lo que denominan *ignorancia racional*: partiendo de la “ventaja comparativa epistémica”, que describí líneas atrás, muchos economistas sostienen que lo más racional es especializarse en aquellos conocimientos, teóricos o prácticos, que sean más relevantes para los oficios de un individuo. Esto lo ejemplificamos con el caso del cirujano, a quien ciertamente le conviene dedicar su tiempo y energías en profundizar sus conocimientos y habilidades en su área de especialidad clínica. Como consecuencia, el cirujano podría no disponer de otros conocimientos que poseen gran relevancia social, como la política o la economía.

En este sentido, Caplan (2007) señala que la ignorancia racional no es problemática en lo que concierne a la esfera privada, mas sí puede serlo en lo que respecta a la esfera pública. Volviendo a nuestro ejemplo, el cirujano puede adoptar una postura xenofóbica antimigración y, si se halla plenamente convencido de esta creencia,

puede votar por candidatos que promuevan tales políticas discriminatorias. Como sostiene Caplan, la ignorancia racional conlleva ciertos peligros en la medida en que los votantes pueden elegir gobernantes cuyas plataformas o programas políticos resulten muy perjudiciales.

Lo crucial es entender que el principal incentivo que tienen los políticos en una democracia es obtener el mayor número de votos. Aristóteles y los retóricos de la Antigüedad ya nos advertían de que el uso de argumentos basados en el *pathos* y el *ethos* son los más persuasivos en el ámbito público. Para lograr empatía con el electorado, el político no buscará cambiar las creencias de su auditorio, sino que, por el contrario, él adaptará su discurso según la demanda de la audiencia, lo cual supone incorporar las creencias más generalizadas entre los votantes (y si adorna este discurso con *tropos* que apelan al *ethos* y *pathos* del público, el éxito persuasivo puede estar garantizado). Del mismo modo, si ciertas creencias económicas *folk* se hallaran muy extendidas entre el electorado, no es de sorprenderse que los políticos las asuman, lo que los motiva a prometer ciertas medidas que están en consonancia con tales creencias. Poco importa si el político realmente cree o no en lo que propone, lo que importa realmente es que las creencias *folk* crean estructuras de incentivos que estimulan a los políticos a aprobar ciertas leyes o programas de gobierno, los cuales podrán tener consecuencias adversas para la economía. Esto se debe a que el político no paga ningún precio por estar equivocado, pues los efectos negativos de las políticas públicas que este ha promovido pueden presentarse a mediano o largo plazo, y para cuando esto ocurra, el político ya se habrá movido a otro lugar en la estructura jerárquica del gobierno.

Tal vez este sea uno de los factores que permite explicar la creciente oleada de populismos autoritarios que han surgido en diversas regiones del planeta recientemente. Este fenómeno ha suscitado reacciones de alarma entre los expertos en ciencia política por el discurso demagógico y por el tipo de políticas públicas que defienden los gobiernos populistas. Pero lo que suelen pasar por alto estos expertos es que el demagogo populista solamente se ha adaptado a las creencias de la gente, y si goza de amplia popularidad es justamente porque ha logrado ganarse su simpatía de

forma muy eficaz. Si algo muestra este fenómeno del populismo autoritario es que son los mismos votantes los que pueden socavar la propia democracia.

Después de todo, el público lego que adopta estas creencias económicas no es un mero espectador que contempla pasivamente los procesos económicos, sino que es un agente económico, y, como tal, puede padecer las consecuencias de las recesiones económicas, como perder su negocio o su empleo. El problema es que si atribuye a los inmigrantes extranjeros el ser la causa del desempleo o adopta “teorías conspirativas” sobre el origen de los males sociales, puede llevar al poder a ciertos gobernantes que establezcan políticas dañinas para la economía en su totalidad. Por ejemplo, la tentativa populista de resucitar las políticas cepalinas de corte mercantilista representa un peligro bastante real para nuestro contexto latinoamericano. Por ello, la importancia de las creencias económicas *folk* no es menor, dadas sus implicaciones para la democracia y para las políticas públicas.

6. CONCLUSIONES

En última instancia, estudiar las concepciones legas es esclarecer cómo los seres humanos conciben el mundo y la realidad social. En particular, investigar a fondo las creencias económicas *folk* permitiría arrojar luz acerca de cómo las personas sin formación económica conciben la dinámica de los mercados y de las actividades productivas, y esto podría ser revelador acerca de cómo reaccionan ante ciertas circunstancias generadas por el desempeño de la economía, como los ciclos económicos. Pero en términos generales, puede aportar importantes elementos para la comprensión de los procesos de justificación epistémica, que ha sido uno de los problemas centrales de la epistemología analítica contemporánea. Estos asuntos también pueden ser de gran importancia para las ciencias cognitivas, sobre todo en lo que toca a la relación entre sesgos cognitivos y creencias. En otras palabras, las investigaciones sobre las creencias *folk* podrían contribuir a la comprensión de ciertos procesos cognitivos que motivan el pensamiento y la acción.

Otro aspecto relevante de este programa de investigación es que podría dar cuenta de algunos aspectos de los procesos de toma de decisión y de las preferencias políticas. Si las creencias pueden motivar ciertas decisiones, resulta claro que el estudio de la economía *folk* posibilitaría la explicación de muchos comportamientos de los individuos en la sociedad. Esto podrá ser de utilidad para investigaciones que pretendan indagar acerca de las causas que motivan las preferencias políticas del electorado. Como se indicó en el apartado anterior, el preocupante avance del populismo autoritario en Latinoamérica y otras regiones del planeta podría estar motivado por esta clase de creencias, entre otros factores.

En suma, si hay un acercamiento cada vez mayor entre las ciencias cognitivas, la epistemología y los estudios políticos, el tema de las concepciones legas podría representar una pequeña pieza que colaboraría en la construcción del rompecabezas.

REFERENCIAS

- Ang, Yuen Yuen. *How China Escaped the Poverty Trap*. Cornell University Press, 2016.
- Ayau Córdón, Manuel F. *El proceso económico. Descripción de los mecanismos espontáneos de la cooperación social*. Editorial Centro de Estudios Económicos Sociales, 2003.
- Baron, Jonathan. *Thinking and Deciding*. Cambridge University Press, 2007.
- Bastiat, Frédéric. “Lo que se ve y lo que no se ve, 1850”. *Obras escogidas*. Madrid: Unión Editorial, 2004. 22-58.
- Bhalla, Surjit S. *The New Wealth of Nations*. Nueva Deli: Simon and Schuster, 2018.
- Bhagwati, Jagdish y Arvind Panagariya. *Why Growth Matters: How Economic Growth in India Reduced Poverty and the Lessons for Other Developing Countries*. Nueva York: Public Affairs, 2013.
- Boyer, Pascal., y Petersen, Michael B. “Folk-economic Beliefs: An Evolutionary Cognitive Model”. *Behavioral and Brain Sciences*. Cambridge University Press, 2018. 1-65. <<https://doi.org/10.1017/s0140525x17001960>>

- Brown, Cecil H. "Folk Botanical Life-forms: Their Universality and Growth". *American Anthropologist* 79.2 (1977): 317-342. <<https://doi.org/10.1525/aa.1977.79.2.02a00080>>
- _____. "Folk Zoological Life-forms: Their Universality and Growth". *American Anthropologist* 81.4 (1979): 791-817. <<https://doi.org/10.1525/AA.1977.79.2.02A00080>>
- Buchanan, James M. "Politics Without Romance: A Sketch of Positive Public Choice Theory and Its Normative Implications". *The Collected Works of James M. Buchanan*. Indianapolis: Liberty Fund, 1999. 45-69.
- Caplan, Bryan. "What Makes People Think Like Economists? Evidence on Economic Cognition from the 'Survey of Americans and Economists on the Economy'". *Journal of Law and Economics* 44.2 (2001): 395-426. <<https://doi.org/10.1086/322812>>
- _____. "Sociotropes, Systemic Bias, and Political Failure: Reflexions on the Survey of Americans and Economists on the Economy". *Social Science Quarterly* 83.2 (2002a): 416-435. <<https://doi.org/10.1111/1540-6237.00092>>
- _____. "Systematic Biased Beliefs About Economics. Robust Evidence on Judgmental Anomalies from the Survey of Americans and Economists on the Economy". *The Economic Journal* 112.479 (2002b): 433-458. <<https://doi.org/10.1111/1468-0297.00041>>
- _____. "How Do Voters form Positive Economic Beliefs? Evidence from the Survey of Americans and Economists on the Economy". *Public Choice* 128.1 (2006): 367-381. <<https://doi.org/10.1007/s11127-006-9026-z>>
- _____. *The Myth of the Rational Voter. Why Democracies Choose Bad Policies*. Princeton University Press, 2007. <<https://doi.org/10.1515/9781400828821>>
- Coleman, William Oliver. *Economics and Its Enemies. Two Centuries of Anti-Economics*. Nueva York: Palgrave Macmillan, 2002.
- Deaton, Angus. *The Great Escape. Health, Wealth, and the Origins of Inequality*. Princeton: Princeton University Press, 2013.
- Descartes, René. *Meditaciones metafísicas (y otros textos)*. Tomo II. Madrid: Gredos, 2011.

- Elster, Jon. *Una introducción a Karl Marx*. Ciudad de México: Siglo XXI, 1991.
- Evans, Jonathan y Keith Frankish. *In Two Minds: Dual Processes and Beyond*. Oxford University Press, 2012.
- Friedman, Milton y Rose Friedman. *Free to Choose: A Personal Statement*. Nueva York: HBJ, 1980.
- Gwartney, James D. et ál. *Common Sense Economics. What Everyone Should Know About Wealth and Economics*. Nueva York: St. Martin's Press, 2016.
- Haidt, Jonathan. *The Righteous Mind. Why Good People Are Divided by Politics and Religion*. Nueva York: Pantheon Books, 2012.
- Hannon, Michael y Jeroen de Ridder. *The Routledge Handbook of Political Epistemology*. Londres: Routledge, 2021.
- Hayes, Patrick J. "The Naive Physics Manifesto". *Expert Systems on Micro-electronic Age*. Ed. Donald Michie. Edinburgh University Press, 1979.
- _____. "The Second Naive Physics Manifesto". *Formal Theories of the Commonsense World*. Eds. Hobbs, Jerry R. y Moore Robert C., Menlo Park CA: SRI International, 1983. 1-36. <<https://doi.org/10.1016/B978-1-4832-1447-4.50010-9>>
- Hazlitt, Henry. *Economics in One Lesson*. Nueva York: Harper & Brothers, 1946.
- Heyne, Paul. "Are Economists Basically Immoral?". *Are Economists Basically Immoral? and Other Essays on Economics, Ethics, and Religion*. Eds. Brennan, Geoffrey y Waterman, A.M.C. Indianápolis: Liberty Fund, 2008. 1-9.
- Hutt, William H. *Politically Impossible...? An Essay on the Supposed Electoral Obstacles Impeding the Economical Analysis into Politics or Why Politicians Do Not Take Economic Advice*. Westerham Kent: Institute of Economic Affairs, 1971.
- Jackendoff, Ray. "The Structure of Language. Why it Matters to Education". Conferencia en Learning and Brain. Cambridge, 5-8 noviembre 2003. <<https://peeps.unet.brandeis.edu/~jackendo/StructureofLanguage1.pdf>>
- Jevons, William Stanley. "Brief Account of a General Mathematical Theory of Political Economy". *Journal of the Royal Statistical Society*, vol. xxix, 1862. 282-287.

- Kahneman, Daniel et ál. *Judgement under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982. <<https://doi.org/10.1126/science.185.4157.1124>>
- Kahneman, Daniel. et ál. *Heuristics and Biases. The Psychology of Intuitive Judgement*. Cambridge University Press, 2002.
- Lal, Deepak. *Poverty and Progress. Realities and Myths about Global Poverty*. Washington: Cato Institute, 2013.
- Mankiw, Gregory N. *Principles of Economics*. Boston: Cengage, 2021.
- Menger, Carl. *Grundsätze der Volkswirtschaftslehre*. Viena: Braumüller, 1871.
- Newcomb, Simon. “The Problem of Economic Education”. *The Quaterly Journal of Economics* 7.4 (1893): 375-399. <<https://doi.org/10.2307/1882282>>
- Niemietz, Kristian. *Socialism: The Failed Idea That Never Dies*. Londres: IEA, 2019.
- Parkin, Michael. *Economics*. Essex: Pearson, 2015.
- Peirce, Charles S. “The Fixation of Belief, 1877”. *Peirce on Signs*. Ed. James Hoopes. University of North Carolina, 1991. 144-159.
- Pinker, Steven. *The Blank Slate. The Modern Denial of Human Nature*. Nueva York: Penguin Books, 2002.
- _____. *Enlightment Now. The Case for Reason, Science, Humanism, and Progress*. Nueva York: Penguin Books, 2018.
- Preston, Carol. “Folk Linguistics”. *Encyclopedia of Language and Linguistics* 2ª ed. Eds. Brown, Keith. Ámsterdam: Elsevier, 2006. 521-533.
- Price, Henry H. “Some Considerations About Belief”. *Proceedings of the Aristotelian Society* 35: Oxford: Oxford University Press, 1934 – 1935. 229-252.
- Rothbard, M. N. “Ludwig von Mises, su esencia”. *Ludwig von Mises. Infatigable luchador contra la economía ficticia*. Ciudad de México: Centro de Estudios en Economía y Educación, A.C. 1983. 20-52.
- Rubin, Paul H. “Folk Economics”. *Southern Economic Journal* 7.1 (2003): 157-171. <<https://doi.org/10.1002/j.2325-8012.2003.tb00561.x>>
- Schettino, Macario. *Introducción a la economía para no economistas*. Pearson Educación de México, 2002.
- _____. *Economía en un día*. Ciudad de México: Paidós, 2015.

- Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of the Nations*, 1776. Indianápolis: Liberty Fund, 1976.
- Smith, Barry y Roberto Casati. “Naive Physics”. *Philosophical Psychology* 7.2 (1994): 227-247. <<https://doi.org/10.1080/09515089408573121>>
- Sowell, Thomas. *Economic Facts and Fallacies*. Nueva York: Perseus Books Group, 2011.
- _____. *Basic Economics*. Nueva York: Perseus Books Group, 2015.
- Von Böhm-Bawerk, Eugen. *Kapital und Kapitalzins*. Innsbruck: Wagner, 1884.
- Walras, Léon. *Éléments d'économie politique pure*. Lausana: Corbaz, 1874.
- Wheelan, Charles. *Naked Economics. Undressing the Dismal Science*. Nueva York: W.W. Norton & Company, 2019.
- Wolpert, Lewis. *The Unnatural Nature of Science*. Cambridge, MA: Harvard University Press, 1992.

FOLK ECONOMICS AND ECONOMIC SCIENCE*

FOLK ECONOMICS AND ECONOMIC SCIENCE*

LUIS ENRIQUE ORTIZ GUTIÉRREZ
Universidad de Guadalajara
Guadalajara, Jalisco.

Luis.Ogutierrez@academicos.Udg.Mx
<https://orcid.org/0000-0003-2680-9230>



ABSTRACT

The development of science has produced a gap between the knowledge accepted by scientific communities and popular beliefs. This subject has worried scientists, considering that many of these beliefs turn out to be untenable. In addition to the problems related to education and scientific dissemination, the study of folk beliefs raises intriguing questions for epistemology and cognitive sciences about the processes of formation and justification of knowledge. This theme also has important repercussions in terms of actions, based on the fact that folk beliefs can motivate wrong decisions. The objective of this article is to analyze the folk beliefs related to the economic field, showing some of its possible causes and some consequences regarding decision making.

Keywords: epistemology; knowledge; belief; folk belief; folk science; economics; epistemic comparative advantage; cognitive bias; morals; politics.

* This article should be cited: Ortiz Gutiérrez, Luis Enrique. "Folk Economics and Economic Science". *Revista Colombiana de Filosofía de la Ciencia* 24.48 (2024): 45-77. <https://doi.org/10.18270/rcfc.4229>

RESUMEN

El desarrollo de la ciencia ha producido una brecha entre el conocimiento aceptado por las comunidades científicas y las creencias populares. Esto ha preocupado a los científicos al considerar que muchas de esas creencias resultan ser insostenibles. Además de los problemas relativos a la educación y la divulgación científica, el estudio de las creencias folk plantea interesantes interrogantes para la epistemología y las ciencias cognitivas acerca de los procesos de formación y justificación del conocimiento. Este tema tiene también importantes repercusiones en cuanto a las acciones, partiendo de que las creencias folk pueden motivar decisiones erróneas. El objetivo del presente artículo es analizar las creencias folk relativas al ámbito económico, mostrando algunas de sus posibles causas y algunas consecuencias en lo tocante a la toma de decisiones.

Palabras clave: epistemología; conocimiento; creencia; creencia folk; ciencia folk; economía; ventaja comparativa epistémica; sesgos cognitivos; moral; política.

“The feeling of believing is a more or less sure indication of there being established in our nature some habit which will determine our actions. Doubt never has such an effect”

Charles Sanders Peirce (1991 149).

1. THE GAP BETWEEN SCIENTIFIC KNOWLEDGE AND “FOLK SCIENCE”

One of the theoretical problems that has occupied specialists from various disciplines is the gap between popular beliefs and scientific knowledge. On one hand, there are various beliefs, adopted by different sectors of the population, about the nature of the universe, living beings, society, language, the mind, etc., that are transmitted through generations. On the other hand, scientists have formulated various theories, subjected to different experimental tests or deductive reasoning, to try to explain

these objects or phenomena, which has led the sciences to progressively distance themselves from these beliefs (Wolpert 1992). In this sense, the gap has grown as the results of scientific research contradict or question these beliefs.

We can mention some examples: people without medical training have multiple beliefs about diseases that do not align with what doctors propose. For instance, based on certain symptoms a person exhibits, people intuitively associate them with a specific illness, such as the common cold, even though these same symptoms can manifest in various conditions, ranging from the common flu to different types of influenza. Thus, laypeople in medicine assume that certain remedies that "cure" the common cold will be effective in all these cases—for example, chicken soup—when in reality, viral respiratory infections do not have a "cure." Instead, doctors prescribe patients analgesics, antihistamines, and antipyretics to manage the symptoms.

Another example is represented by lay linguistic conceptions. Linguist Ray Jackendoff (2003, 3–4) lists a series of beliefs about language: 1) "There is a correct way to speak a language" (in contrast, linguists reject the notion of a "correct" use of language and point out that the multiple uses of language are due to dialectal, sociolectal, diachronic, etc., variations); 2) "Language is deteriorating due to improper usage" (this belief ignores linguistic change in both lexicon and grammar, which is inevitable); 3) "Dictionaries tell us the correct way to speak" (but dictionaries are merely incomplete and partial inventories of certain lexical samples, many of which belong to dialectal variants); 4) "Written language is the language" (however, historically, the appearance of writing is late, and many present languages do not have their own writing system). Generally speaking, these beliefs of speakers assume that there is a rigid normative model of language, and any use that does not correspond to this model will be considered "incorrect."

In this way, there is a wide variety of beliefs about nature, health, language, life, society, etc. that contrast with the principles accepted by scientists. Of course, the existence of such beliefs does not represent a problem. As the American philosopher Charles S. Peirce (1991, 144-ss) pointed out, beliefs represent an intermediate point between doubt and scientific knowledge. However, beliefs are not merely

provisional knowledge with which individuals form a worldview that provides them with a certain degree of certainty; they also produce motivations for action. Remaining in a state of doubt keeps the individual inactive, and therefore, a certain degree of certainty provided by belief is required to act in the world.

In any case, the gap between popular beliefs and scientific statements has increased significantly. It is possible that this gap is due to certain divergences in the processes of knowledge formation: while scientific theories depend on complex processes of cognitive elaboration, popular belief systems seem to be produced by individuals' intuitions and cultural traditions, among other possible factors. This means that understanding scientific theories requires theoretical training and increasing specialization, which could explain why the gap continues to widen.

For some years now, this topic has been developed in various studies within certain scientific disciplines. In cognitive psychology, for example, Steven Pinker (2002, 195–196) presents a list of certain "cognitive faculties" that allow humans to form intuitive knowledge with which they attempt to understand and, consequently, act in the surrounding world. This list, which the author acknowledges as provisional, includes *intuitive physics*, *intuitive biology*, *intuitive linguistics*, *intuitive engineering*, *intuitive psychology*, *intuitive economics*, etc.

Other research has delved deeper into the phenomenon. Under the adjective 'folk,' popular beliefs regarding certain objects of study within specific disciplines have been grouped: thus, we speak of "folk physics" (Hayes 1979, 1983; Smith 1994), "folk biology" (Brown 1977, 1979), "folk psychology" (Pinker 2002), "folk linguistics" (Preston 2006), etc. The aim of these studies is to identify folk beliefs and contrast them with scientific statements to highlight their differences. Some of the purposes include attempting to develop strategies for science dissemination, the teaching of scientific knowledge, and advising on science and technology legislation, among others.

The aim of this paper is to analyze folk beliefs in a very specific discipline: economics. In particular, it will focus on identifying some of the most common folk economic beliefs and presenting some hypotheses about their possible causes. Of course, such hypotheses will require extensive empirical corroboration, so this work

represents the outline of a research program that may have important implications for epistemology and the philosophy of science, not to mention the possible formulation of guidelines for the development of educational policies aimed at promoting and disseminating scientific knowledge.

Later, we will examine some implications that folk economics might have for politics. If beliefs, as Peirce asserts, establish motivations for action, then their study can be of paramount importance for describing and explaining individuals' behaviors. This becomes highly relevant for modern democracies, as public policies are decided based on voters' participation in electoral processes. But if voters make political decisions based on mistaken beliefs, what consequences could this have for democracy?

2. DEFINITIONS

In epistemology, philosophers have long debated the relationship between knowledge and beliefs. This has led to reflection on how to define these notions more clearly and precisely. Of course, as often happens in philosophy, all definitions are problematic without exception, and what will be presented below is no exception.

In the first instance, it is necessary to clarify that the distinction between *belief* and *knowledge*, or between *expert* and *lay audiences*, simply serves the purpose of distinguishing certain groups of individuals and specific types of knowledge. Referring to a *lay audience* does not, in any way, characterize certain people as 'ignorant' or 'people with low intellectual capacity,' because, in some way, *we are all laypersons*: it is impossible to have full knowledge of every theoretical discipline, and, in reality, people specialize in certain areas of knowledge, whether theoretical or practical, according to their individual capacities, preferences, or interests. Thus, a mathematician may be a recognized expert in their field but know nothing about basic biology, and a biologist may be an authority on population genetics but incompetent at changing a car tire. This clarification is important to prevent the use of these terms from causing confusion, misunderstandings, or misconceptions.

The definition of belief proposed by the philosopher H. H. Price (1935 232), which is a reformulation of the definition by the also-philosopher Cook Wilson, will be adopted for the purpose of this text. It states that a belief consists of the following:

- Given a subject S, it is said that S believes in the proposition p
- if S is aware of or considers p
- and S assents to or adopts p .

To illustrate this definition, let's suppose that Betty does not know where her cat is. Betty considers or is aware of a series of propositions: "the cat is in the pantry" (A), "the cat is in the dining room" (C), and "the cat is in the garden" (J), as possible descriptions of her pet's location. If Betty assents to or adopts C, thereby discarding A and J, then it can be said that "Betty believes that C" ("Betty believes that the cat is in the dining room"), that is, Betty assumes that C is true, thus excluding the alternative propositions A and J. In other words, *a belief is a proposition that is considered and adopted as true by a subject*.

The advantage of this definition by Price and Wilson lies precisely in its foundation on the opposition between belief and knowledge. In our example, Betty's belief in C implies that she leans towards or assents to C, meaning she considers it true and therefore rejects the other propositions. However, C could be false. Thus, the condition of any belief is that there is always the possibility that the proposition is false, even though the subject adopts it as true. In contrast, talking about knowledge implies that the proposition is considered, first and foremost, true. This is where epistemological controversies arise, because the criteria of truth and justification that have been proposed are subject to debate among philosophers.

Therefore, the following criteria are proposed to provisionally characterize knowledge: a true belief will be considered knowledge if a community of experts has a set of reasons or certain evidence to validate this proposition, thereby discarding other alternative propositions. Evidently, this does not resolve the problem of epistemic justification, but it can serve as a starting point for further analysis.

Let us now define what a folk belief is. Our approach follows the proposal by Price and Wilson: we assert that *S* holds a folk belief if *S* adopts a proposition *p*, and *p* is in opposition to a proposition *q*, which refers to the same fact about the world. However, unlike *p*, *q* is supported by sufficient reasons to be considered knowledge by experts in the relevant field. This can be expressed as follows:

The belief in *p* is a folk belief if:

- *S* believes that *p*.
- A community of experts believes that *q*.
- *p* is in opposition to *q*.

The following example will illustrate this concept. Pedro firmly believes that the Earth is flat (*T*), which is in opposition to the proposition that the Earth is an oblate spheroid (*E*) held by astronomy. Therefore, Pedro's belief that the Earth is flat is a folk belief if:

- Pedro believes that *T*.
- The community of astronomers believes that *E*.
- *T* is in opposition to *E*.

Both *T* and *E* describe the same fact about the world, in this case, the shape of our planet. However, *E* is considered knowledge insofar as astronomers have provided certain scientific evidence to support it: Eratosthenes' experiments comparing shadows at different latitudes, the apparent sinking of ships as they move away over the horizon, the Earth's shadow cast on the Moon during a lunar eclipse, photographs taken by artificial satellites, etc. Thus, the opposition between *T* and *E* lies in the fact that *E* is accepted by astronomers as knowledge, whereas *T* is not.

However, the distinctions established so far are insufficient, as they do not provide a criterion to distinguish between possibly false beliefs held by an individual

and folk beliefs. If Pedro believes that his grandfather's soul transmigrated to his dog, it may be a false belief, but it would not necessarily be a folk belief. However, if a community of individuals adopts the belief that people's souls transmigrate to pets, this would indeed constitute a folk belief. In this sense, folk beliefs have an important distinguishing feature: their social character.

Folk beliefs are shared by a given community, which we can refer to as the "lay public." The lay public is defined as those who lack training in a specific area or discipline of knowledge. In our example, T would be a folk belief insofar as T is agreed to or adopted as true by a certain community that lacks training in astronomy.

The definition of a folk belief, then, would be as follows: *a proposition accepted or adopted by the lay public that stands in opposition to other propositions recognized as knowledge by experts in a specific field or discipline.*

Thus, the so-called "folk sciences" would be comprised of: 1) a set of beliefs that describe certain facts that are alternatively explained by specific scientific theories; 2) these beliefs contradict the explanations provided by those theories; and 3) there is a community of laypeople who have adopted such sets of beliefs. Seen this way, calling the flat earth view a 'theory' is an excessive concession, as it is not truly a scientific theory but, at best, a belief typical of "folk physics." The same could be said of creationism and many other beliefs that are unjustifiably granted the status of 'theories.'

Once the basic concepts of our analysis are clarified, we will now review the most common folk economic beliefs.

3. FOLK ECONOMICS VERSUS ECONOMIC SCIENCE

Since the emergence of modern economics, there have been several attempts to disseminate economic science to the lay public (Bastiat 2004; Newcomb 1893; Hazlitt 1946; Friedman & Friedman 1980; Sowell 2015; Schettino 2002, 2015). Despite these various efforts, the dissemination of economics has met with very little success. This is not surprising, as similar outcomes have occurred with popularization programs in physics, biology, linguistics, and other scientific disciplines.

This has led some theorists to investigate the reasons behind the failure of economic dissemination, encountering the gap between lay conceptions and scientific knowledge, in this case, economic theories. Bryan Caplan (2001, 2002a, 2002b, 2006, 2007), for example, has pointed out that people without economic training exhibit certain cognitive biases when attempting to explain the functioning of the economic system, although he does not use the term folk belief. However, more recent research (Rubin, Boyer, and Petersen) has proposed the term 'folk-economic beliefs' and shows results that align with Caplan's earlier studies. Thus, there is already a body of research on folk-economic beliefs. The question now is to determine what these economic beliefs consist of, which form what we could designate as 'folk economics.'

Clearly, addressing this issue goes beyond merely listing the beliefs held by laypeople regarding economics. Therefore, it is essential to specify the defining characteristics of these beliefs by examining their opposition to the propositions accepted as knowledge in economic science.

First, it is necessary to acknowledge that in economic science, there are numerous debates regarding specific issues (for example, proposals to address externalities or specific aspects of monetary policies). However, there are also certain basic principles broadly accepted by the majority of researchers within the mainstream of economic science (Ayau 2003; Parkin 2015; Gwartney et al. 2016; Sowell 2015; Mankiw 2021). We can highlight the following:

- A1) A country's wealth does not derive from the amount of money or precious metals it possesses but from the productivity of its inhabitants.
- B1) The pursuit of self-interest contributes to general welfare. The market represents a positive-sum game (in an economic exchange, both parties mutually benefit; it's a win-win situation).
- C1) For both individuals and nations, it is optimal to specialize in producing goods for which they have the best capabilities or resources and to purchase other goods from individuals or countries, even if those goods could also be produced domestically (principle of comparative advantage).

- D1) The price system provides information about supply and demand relationships and allows for the coordination of economic activities in the market without the need for a central planning authority.
- E1) The value of goods and services is subjective. It is not an objective property of things but depends solely on the desires and preferences of individuals.
- F1) The value of goods and services is marginal and relative. It is not an absolute property but depends on previous possession, use, or consumption of those same goods.

These are some of the basic propositions accepted as knowledge by consensus among the majority of economists. Classical economics established principles A1 through D1, while neoclassical economics included E1 and F1. From these propositions, we can derive some theoretical implications. Firstly, economic science starts with the primacy of the individual, placing them at the center of economic processes. Individuals seek to satisfy their subjective desires and preferences, which creates incentives for consumers, intermediaries, and producers to interact with one another. Given the relative scarcity of resources that have alternative uses, it is essential that these resources be allocated in the most efficient way possible. For example, cotton can be used to produce clothing, ropes, medical supplies, biofuels, etc. How much cotton will be sent to each of the industries that produce these materials and how economic agents will coordinate to carry out these activities is the great challenge of any economic system, and it is also the central object of study in economic science.

The goal of economic exchanges, then, is to achieve the most efficient allocation of scarce resources to produce goods and services, which is accomplished through the price system in market economies. This system acts as a means of conveying information about consumer demand and the availability of scarce resources (supply) to meet that demand. In this sense, prices are determined by the number of additional ("marginal") units of goods that can be produced by suppliers and purchased by consumers. Similarly, prices create incentives that encourage suppliers to produce certain goods and compel consumers to adjust their spending. In very

general terms, this is how the problem of allocating scarce resources with alternative uses and coordinating economic activities is addressed.

In contrast to these principles, some of the most widespread folk economic beliefs are the following:

- A2) A country's wealth is derived from the amount of money or precious metals it possesses.
- B2) The pursuit of self-interest only contributes to personal enrichment. The market represents a zero-sum game (in an economic exchange, one party benefits at the expense of the other; it is a win-lose situation).
- C2) The best thing for a country is to reduce its imports and increase its exports. The national economy should aim for self-sufficiency (the principle of autarky).
- D2) The price system reflects the greed of intermediaries, and government intervention through price controls is necessary.
- E2) The value of goods and services is an intrinsic objective property, produced by the amount of labor invested in them.
- F2) The value of goods and services is absolute; it does not depend on circumstances or the values of other goods.

As can be seen, these folk beliefs are in clear opposition to the aforementioned economic principles. Specifically, the oppositions are as follows:

- The opposition between A1 and A2 lies primarily in a semantic difference. The definition of the term "wealth" in economic science differs from its colloquial use, as can be seen in the online dictionary entry of the Real Academia de la Lengua Española:
 1. f. Abundance of valuable goods and precious things.
 2. f. Abundance of excellent qualities or attributes.

3. f. Relative abundance of anything. Alcoholic wealth, wealth of minerals, wealth of vocabulary.

Of the three definitions, only the first seems to relate to the economic conception. It defines 'wealth' as the possession of valuable goods and 'precious things,' clearly corresponding to belief A2. In economics, since Adam Smith (in his classic *Wealth of Nations*, 1976), wealth refers to the goods and services produced to satisfy individual desires. Although the amount of money an individual has allows them to obtain more goods to satisfy their desires, money itself does not constitute wealth; it is merely a medium of exchange to obtain these goods. In fact, if a country's wealth were defined solely by the amount of money it possesses, simply printing more currency would make that country rich. However, as economists have repeatedly warned, printing large amounts of money only leads to inflationary pressures that negatively affect national economies.

Thus, it is not the amount of money or 'precious things' that characterizes wealth, but rather the production of goods and services. As Smith already emphasized, what determines the increase in wealth is productivity.

In contrast, belief A2 assumes that wealth is something pre-existing: it is conceived as a finite set of goods (precious metals, agricultural products, oil, etc.) that has existed since the beginning of time, with individuals or social groups fighting each other to obtain a portion of it (Pinker 2018, 80). As we will see, the following folk beliefs are closely related to this one, because just as principle A1 is the foundation of modern economic theory, principle A2 represents what we might call the "mother economic belief."

- The opposition between B1 and B2 is closely related to the aforementioned concepts. The idea that the market is a zero-sum game (Rubin 2003, 157ff; Sowell 2011, 3ff) can be illustrated with Milton Friedman's metaphor of the pie. If wealth is imagined as a pie of fixed size—that is, wealth always represents a fixed amount (A2)—it is conceived that if an individual or group of individuals receives a larger slice of the

pie, the rest of the population will receive proportionally smaller slices. Hence, it is inferred that the main cause of social problems stems from an “unfair distribution” of the pie.

The reality is that the market is not fixed, nor is the economy a zero-sum game; it is not a pie, but a machine that massively produces pies. If economic exchanges were zero-sum games, the parties involved simply would not participate in the exchanges. Moreover, if in a market economy the increase in wealth came at the expense of large sectors of the population, this would be empirically shown by a correlation between economic growth and increased poverty. However, the spectacular economic growth and drastic reduction of poverty in China—since 1978—and in India—since 1991—show a correlation in the opposite direction (Bhagwati & Panagariya, 2013; Ang, 2016; Bhalla, 2018). These cases demonstrate that the improvement in living standards for large sectors of the population and economic liberalization go hand in hand (Deaton 2013; Lal 2013). Therefore, the belief that the market is a zero-sum game is simply false.

On the other hand, belief B2 stems from the presumption that only selfless acts, such as altruistic actions, can be socially beneficial. From this, it follows that actions aimed at satisfying personal interests can only benefit the individuals who perform them. What lies behind this belief is a certain moral valuation, as it implicitly condemns the pursuit of personal satisfaction: if a person becomes wealthy, they are seen as greedy and selfish. In a more sophisticated form, some theorists have accused economics of promoting selfishness (Heyne 2008, 1ff; Coleman 2002, 136ff). In summary, the problem with B2 is that it excludes the possibility that the pursuit of personal satisfaction—which, according to economists, is the primary motivation for economic exchange—can be socially beneficial (Ayau 2003, 2ff).

The opposition between C1 and C2 presents a peculiarity. Individuals often apply the principle of comparative advantage (Gwartney

et al. 2016, 36) to their personal activities (for example, although a person may be very skilled at cleaning their home, they might choose to hire someone else to do it, as this allows them to save time that can be dedicated to more important or productive activities) and may have no problem accepting this principle theoretically in personal matters but reject it when it pertains to public issues, such as the national economy. In other words, laypeople may accept C1 when it comes to their private interests but accept C2 in the public sphere. An example of this is the belief that imports harm the national economy or that national economies should strive for self-sufficiency. However, as economists point out (Mankiw 2021, 52), the costs of engaging in many varied economic activities exceed the potential benefits, as resources are being allocated to produce goods that are more costly to produce locally. Therefore, the most beneficial strategy for an economy would be to focus its resources on producing goods that involve lower costs or generate more profits, i.e., to take advantage of a country's comparative advantages. Evidence that the lay public understands comparative advantage can be seen in the old adage "Jack of all trades, master of none"; the issue is that they seem not to apply it beyond their private sphere.

The opposition between D1 and D2 stems not only from a limited and partial perspective on how the price system works, as implied in D2, but also from another assumption: that government intervention is the best way to benefit society. From the layperson's perspective, if economic exchanges are zero-sum games and the pursuit of self-interest involves selfish actions, it is assumed that the state is a third party that prevents abuse or exploitation by certain economic agents, such as intermediaries (Sowell 2011; Wheelan 2019). The problem is that the layperson assumes that this third party has all the necessary information to achieve the efficient allocation of goods and services and, at the same time, represents the moral authority to presumably achieve general welfare—in other words, it presupposes that the state is omnis-

cient and altruistic. However, according to economists, price controls and other interventionist measures distort the information generated in economic exchanges through the price system (Sowell 2015, 92; Gwartney et al. 2016, 186). Moreover, individuals working in government institutions also seek to satisfy their personal interests, just like any economic agent (Buchanan). Therefore, it is difficult to uphold the assumed altruism of the rulers.

Regarding the subjective conception of value (E1), it is clearly incompatible with the objective conception (E2). The latter is not only deeply rooted in the lay public but was also accepted by classical economists (Adam Smith, David Ricardo, François Quesnay, Thomas Malthus, John Stuart Mill, etc.) because they lacked an alternative theory at the time. This was largely because classical economists studied economic phenomena in terms of "classes" of goods and values, not individual preferences (Rothbard 1983, 20–27). It wasn't until the marginalist revolution, initiated by William S. Jevons (1862) and Leon Walras (1874) and extensively developed by the Austrians Carl Menger (1871) and Eugen von Böhm-Bawerk (1884), that the subjective conception of value and marginal analysis were fully developed, later recognized as the theoretical pillars of neoclassical economics.

The lay public, however, assumes that there are "objective valuations" of goods or services, meaning that objects possess inherent value independent of individual preferences and desires. But, as Menger pointed out, each individual has a "preference list" by which consumers subjectively rank goods according to the degree of satisfaction (utility) they derive from them. Comparing the "preference lists" of different individuals shows notable variations regarding which kinds of goods are included in each person's list; even if some goods appear in the lists of two individuals, they may be ranked differently based on the expected utility.

- However, valuations are not only subjective and differ from one individual to another (E1), but they can also change within the same indivi-

dual under various circumstances (F1): a consumer may value a lemon ice cream more in Cancun than in the frigid zones of Canada; they may value coffee more in the morning than at night; they may value savings more in adulthood than in youth, etc. Consequently, the value of a good or service will depend on how many marginal units are available and how many of these the consumer is willing to acquire, given certain circumstances. In contrast, the lay public believes that goods have an absolute value (F2): coffee or lemon ice cream is always valuable at all times and in all places.

The key point here is that marginal analysis allows for the explanation of multiple economic phenomena. For example, some external observers might think that a restaurant could significantly improve its service if the owner hired more staff. However, the owner must consider whether hiring more workers contributes to the productivity of the business. If the marginal utility of a new employee exceeds their marginal cost, then it will be worth paying an extra salary; if not, the most prudent decision for the business is not to increase the workforce. After all, the number of customers and the workload will vary at different times of the day and on different days of the week, factors that the employer takes into account before making such decisions. In short, understanding particular situations in marginal terms allows for a more precise explanation of phenomena rather than adopting an absolute and rigid approach, characteristic of lay conceptions of economics.

In this way, we present the set of oppositions between economic science and the beliefs that make up folk economics. It is worth noting that some of these beliefs historically preceded the emergence of economics as a scientific discipline, just as essentialist and vitalist beliefs preceded modern evolutionary biology or geocentric conceptions preceded modern astronomy. For example, beliefs A2 and B2 were held by mercantilists (16th to 18th centuries), whose ideas significantly influenced anti-free trade policies at that time. Mercantilists argued that a nation's wealth consisted of its inventory of precious metals, so they recommended that a country reduce its imports and increase its exports, as buying more imported products would mean

a reduction of assets in the national treasury. Much of the theoretical effort of classical economists was aimed at combating this doctrine. Paradoxically, mercantilist prescriptions were resurrected by the Economic Commission for Latin America and the Caribbean (CEPAL) in the 1950s, evidencing the persistence of certain erroneous beliefs. This issue will be addressed later.

4. POSSIBLE EXPLANATIONS FOR FOLK ECONOMIC BELIEFS

Based on the previous discussion, the question now arises as to why laypeople adopt such beliefs. As possible answers to this question, I will present the following hypotheses along with the respective arguments that aim to show why I believe these hypotheses could explain this phenomenon:

H1) *Folk economic beliefs are caused by the limited perceptions of phenomena by the lay public.* People tend to perceive and interpret phenomena based on their own circumstances. As with other sciences, the phenomena described and attempted to be explained by economic science require a certain degree of abstraction, as they involve complex systemic processes that transcend the local sphere in which individuals operate and are difficult to understand without the theoretical tools of economic science. Goods that are so common and necessary for individuals, such as clothing, appliances, mobile phones, and food, are the result of the coordination of thousands or millions of economic agents who do not know each other and will probably never meet. Such coordination has not been planned or ordered by any feudal lord or central committee but rather emerges through a myriad of individual interactions. Therefore, many fundamental economic principles are counterintuitive to laypeople and challenge their conceptions of how society works.

Perhaps this helps explain the popularity of so-called “conspiracy theories”—another example of the misuse of the term “theory”—which claim that the world economy is controlled by ultra-secret cabals. From a lay perspective, it is inconceivable that economic events depend on systemic processes that emerge from countless individual exchanges, leading to the assumption that there must be some kind of

hidden control coordinating the vast economic activity for some nefarious purpose. Underlying this way of thinking is the idea that all events result from intentional acts, a common feature of many naive conceptions, not just those related to economics. In the past, it was commonly believed that natural phenomena like rain, earthquakes, and eclipses were deliberately caused by gods or other supernatural beings. The problem with “conspiracy theories” is that they are based on false assumptions: 1) that the secret conspirators possess all the knowledge circulating in the economy (which is impossible, given the numerous factors influencing economic events), and 2) that chance does not exist in the world. Interestingly, economic science, over its nearly two hundred years of existence, has been gradually developing certain theoretical models and methodological tools to attempt to explain the functioning of the economy as precisely as possible, while conspiracy “theorists” already have all the answers. In any case, these pseudotheories may have their origins in certain conceptions of folk economics.

The classical economist Frédéric Bastiat (2004) had already shown that many confusions in economic matters stem from individuals' limited perspectives, as they tend to consider only the benefits of certain decisions—“what is seen”—while overlooking their costs or ignoring the unintended consequences of certain actions—“what is not seen.” For example, Bastiat questions the belief that taxes are inherently beneficial to the economy, arguing that while certain sectors may benefit from government spending, the invisible costs imposed by taxation can exceed these benefits, potentially impacting the economy as a whole. Thus, folk beliefs may arise from a limited and partial understanding of how the economy functions.

H2) Folk economic beliefs are caused by moral valuations. It is common for people to make value judgments before understanding certain phenomena. This tendency to confuse explanations with valuations was already observed by the philosopher René Descartes (2011, Fourth Meditation), who argued that a frequent cause of error is that the will precedes the understanding: we accept something as true because it is pleasing (or morally desirable) and, conversely, reject something as false because it is displeasing (or morally undesirable). The issue is that the truth or fal-

sity of propositions is indifferent to moral valuations.¹ For example, price increases often lead to complaints from consumers, who accuse intermediaries of being abusive and seeking to "profit" from people's needs. Suppose a person wants to rent an apartment in City X but finds that prices are very high in certain areas. They might accuse real estate companies of being "greedy" and charging "excessive" prices. But if prices are lower in some parts of the city, does that mean greed is "lesser" in those areas? The fact that property prices (for sale or rent) are higher in neighborhood 1 than in neighborhood 2 indicates that, on the one hand, such properties are relatively scarce, and on the other hand, there is a higher demand for houses and apartments in neighborhood 1 than in neighborhood 2. Thus, it is not difficult to understand that property prices are comparatively lower in dangerous neighborhoods (such as might be the case in neighborhood 2) than in other parts of the city. A lack of basic understanding of price dynamics—specifically, the factors affecting supply and demand—can lead to erroneous explanations loaded with hasty moral valuations.

Previously, it was noted that many beliefs give rise to certain moralistic judgments, although the causal relationship could be the reverse: lay beliefs may result from this confusion between explanation and moral valuation. If this is the case, it is possible that folk economic beliefs stem from certain value systems previously assumed by individuals, values that could, in turn, be inculcated by tradition or religious upbringing—for example, the famous biblical saying: "Truly, I tell you, it is hard for someone who is rich to enter the kingdom of heaven [...] it is easier for a camel to go through the eye of a needle than for someone who is rich to enter the kingdom of heaven" (Matthew 19:23–30). Therefore, it is possible that certain moral valuations give rise to certain lay economic conceptions.

H3) *Folk economic beliefs are caused by individuals' cognitive specialization in certain activities or areas of knowledge.* This hypothesis is based on two fundamental

¹ For more on the differences between morality and science, and more specifically, between morality and economics, see Coleman (2002, 119–131).

facts: 1) individuals possess different capacities and skills (whether innate or acquired), and 2) individuals have limited time and energy, so they decide to invest their time and energy only in developing those capacities and skills (Ayau 2003, 2ff). Consider the case of a surgeon who has become a prestigious physician in his specialty. He has dedicated much time and effort to acquiring knowledge in his discipline, practiced it for years, developed certain skills, and must continuously stay informed about the latest advances in his field; this represents an "epistemic comparative advantage." This leaves him little time to inform himself about other areas of knowledge or activities, such as political or economic theory. For the surgeon, it would be very costly to dedicate time to studying these or other areas unrelated (directly) to his activity, as it would involve an investment of time and effort that could be better spent on his own discipline. Therefore, the surgeon might adopt folk economic beliefs, such as believing that the best model for the country is one based on autarky. Similarly, an economist might form folk medical beliefs—for example, believing in dubious remedies such as chlorine dioxide "curing" COVID-19—for basically the same reason: cognitive specialization in his area of knowledge. In short, we cannot know everything given our human limitations, and consequently, we focus on learning the knowledge relevant to our activities (this is the "epistemic comparative advantage"). This may be a cause of the adoption of such folk economic beliefs.

H4) *Folk economic beliefs are caused by cognitive biases.* Recent studies in cognitive psychology (Kahneman et al. 1982; Kahneman et al. 2002; Baron 2007; Evans & Frankish 2012) have opened a vast field of research, extending to medicine, politics, and economics, regarding systematic errors individuals make in their reasoning processes that can be linked to decision-making. According to these investigations, humans proceed cognitively in two ways: 1) an intuitive process, consisting of immediate reactions that form judgments under certain circumstances, and 2) a rational, complex, and abstract process, requiring greater cognitive effort to reflect on more general concepts and situations. These "mental shortcuts," called cognitive biases, result from applying intuitions to certain abstract and complex objects. One case is the "availability heuristic" (Pinker 2018, 41-42), which involves making a judgment based on the first thing that comes to the subject's mind. This can be seen

in people who consider traveling by road to be safer than flying simply because a recent news report covered an airplane crash. However, statistics indicate that road accidents are more frequent and that air travel is actually safer.

Caplan's studies (2002b) suggest that lay conceptions in economics could be explained in terms of cognitive biases. An example could be the "pessimistic bias," where individuals believe that the economy's performance has worsened year after year. Laypeople might arrive at this conclusion because a major company went bankrupt, prices have increased, or close acquaintances have lost their jobs, even though statistics may indicate the opposite. Of course, these individuals may dismiss the statistics, likely because they trust their own intuitions more. Thus, cognitive biases could provide a causal model explaining the adoption of folk economic beliefs.

H5) Folk economic beliefs are caused by an asymmetry between private and public decisions. Norwegian sociologist Jon Elster (1991, 40) explores the Marxist idea that throughout history, there have been social contradictions, pointing out that individuals can hold contradictory beliefs and desires. For example, a capitalist might want their workers to have low wages because it benefits them personally, while simultaneously wanting other capitalists to pay high wages to their workers so that these workers can afford to buy their goods or services. Another example is that some businesspeople may advocate for free market rhetoric but contradict it in practice when they lobby in political circles to pass laws or policies that benefit their own interests, such as demanding tariffs to protect their businesses from foreign competition or requesting government subsidies (Hutt 1971; Buchanan 1999).

However, this asymmetry between private and public decisions is not exclusive to intellectual elites or powerful business circles, as all economic agents can experience a conflict between what they consider best for themselves and what they consider best for others. In the previous section, I mentioned that comparative advantage is often accepted by individuals when it comes to their personal activities but may be rejected when it pertains to the *Res Publica*. Another example of this asymmetry could be opinions regarding income taxes. Many people believe that progressive taxation should be implemented to "redistribute wealth," so that the wealthier pay a higher rate, but these same people often complain bitterly when their

own salaries are significantly reduced by income tax withholdings. These examples suggest that there might be a certain pattern that induces individuals to fall into certain contradictory behaviors, leading them to conceive economic phenomena one way in their private sphere and another way in the public sphere. Thus, the asymmetry between public and private decisions could be the cause of these beliefs.

The previous hypotheses could be complementary. H1 and H3 complement each other in that the partial observation of economic phenomena points to limited or nonexistent knowledge, which could result from the "epistemic comparative advantage." Both are also compatible with H2, as partial observation can be interwoven with moral valuations, as suggested by the example about the greed of intermediaries. Likewise, it is possible that a layperson in economics avoids mixing explanations with moral valuations in their field of expertise but does not avoid it in fields that are unfamiliar; for instance, our surgeon may evaluate a certain treatment based on its effectiveness rather than its potential moral value.

Regarding H5, it could also complement the previous hypotheses: moral valuations may be linked to how individuals perceive and evaluate the public sphere,² but do not apply them to their own decisions, leading to performative contradictions (for example, a person may argue that it is best to buy from domestic producers to support them against foreign competition, but at the same time, this person may frequently buy imported products). In this way, partial perception, the confusion between moral valuations and explanations, specialization in certain areas of knowledge, and the asymmetry between the private and public spheres may be closely related.

The hypotheses mentioned above also complement H4 to the extent that cognitive biases may point to certain mechanisms in cognition that could be the ultimate cause of these types of beliefs and the resulting (sometimes inconsistent)

² As possible support for this hypothesis, Pinker (2002, 240) notes that individuals can "turn on and off" their moral feelings and thus judge certain acts either from the perspective of preference or from the perspective of value (that is, some acts are judged as "moral," while others are seen as a mere matter of taste). This can be relevant to the problem we are discussing here.

decisions. Evidently, whether the hypotheses truly complement or contradict each other will be determined through empirical investigation.

However, there is a theoretical framework developed by evolutionary psychology that might unify the previously discussed hypotheses (Pinker 2002; Rubin 2003; Boyer & Peterson 2018). This theory essentially holds that there is possibly a mismatch between the capabilities of the human brain and the specific conditions of contemporary societies. The theory posits that our brains evolved in an adaptive environment very different from the one we live in today.

The main argument of this theory hinges on the temporal dimension. Our species has been inhabiting this planet for approximately 350,000 to 300,000 years. The oldest known civilization dates back about 9,000 years, and the current industrialized commercial societies originated only about 200 years ago. The human brain could not have evolved to respond to these new conditions, as evolutionary changes require long periods of time. In other words, we still have hunter-gatherer brains, even though we live in more complex and advanced societies compared to those of our ancestors. But what characterized the conditions of early humans?

In those times, social interactions may have been founded on reciprocal altruism, encapsulated by the famous saying, "I'll scratch your back if you scratch mine." Given the extreme conditions faced by early *Homo sapiens*, where food sources were very scarce, economic interactions could have been guided by this kind of altruism, which compelled all community members to share what they hunted or gathered. If any member accumulated more than others, it would result in a gain at the expense of the rest of the group—in other words, it would be a zero-sum game. If this were the case, selfish behavior might have warranted social sanctions; for example, tribe members might refuse to cooperate in the future with those who behaved selfishly, thereby endangering their own survival. Morality based on reciprocal altruism is essentially intentional: it requires community members to know the intentions of their comrades to identify whether they are willing to cooperate. Consequently, reciprocal altruism can only function in small societies, and it is not surprising that it persists today in small social groups, such as families.

However, economic interactions in our contemporary society are very different because they are based on trade and industrial production, which create conditions conducive to positive-sum games. As we highlighted, the coordination of economic activities in large modern societies involves interactions among thousands or millions of individuals who do not know each other. Under these circumstances, formal legal systems and impersonal institutional frameworks have been developed to regulate the establishment of contracts and ensure the legal security of economic agents. This is why the reciprocal altruism of primitive tribes simply does not scale up—it is impossible to track the intentions of thousands or millions of individual agents participating in these vast systemic processes. Characterizing modern societies as "one big family" may be a rhetorically appealing metaphor, but it is entirely inapplicable in descriptive and explanatory terms.

Consequently, if this theoretical framework is correct, it is possible that the cause of folk economic beliefs derives from the mismatch between the interaction modes of our ancestors, which are encoded in the human brain, and the current socioeconomic reality. As evolutionary psychologists establish, there are two types of explanations in evolutionary terms: proximate (which answer the question "How does X work?") and ultimate (which answer the question "Why does X exist?"). Hypotheses H1, H2, and H3 could be understood as proximate explanations since they account for specific factors that cause the phenomenon, while H4 could be considered an ultimate explanation, as it might be based on a general evolutionary mechanism that predisposes the formation of folk economic beliefs.

Similarly, this theory may help to understand the formation of certain ethical doctrines associated with these kinds of beliefs. According to Heyne (2008), our moral intuitions are characterized by what he calls "face-to-face morals," which align with the reciprocal altruism described earlier. In this moral system, as indicated, the emphasis is on the intentions of moral agents, and selfish behaviors are condemned. In this sense, the ethical doctrines proposed by many religions and by ancient philosophers like Plato, Aristotle, or Seneca, which explicitly condemn "greed," "avarice," and other behaviors, may have an evolutionary origin. Perhaps the difference between these doctrines and folk beliefs lies in the former being formulated in a more systematic and

abstract manner. It follows that the mismatch between our hunter-gatherer brains and the conditions of modern market economies could explain the persistence of "face-to-face morals" and the philosophical doctrines inspired by them.

However, this theory is still very speculative (Niemietz 2019, 286). As is often the case with theoretical frameworks in evolutionary psychology, the problem lies in the difficulties of finding evidence about the phylogeny of cognitive processes and the living conditions of our primitive ancestors. Despite this, I believe this theory deserves consideration, as it is not ruled out that data supporting it may be found in the future.

In those times, social interactions may have been founded on reciprocal altruism, encapsulated by the famous saying, "I'll scratch your back if you scratch mine." Given the extreme conditions faced by early *Homo sapiens*, where food sources were very scarce, economic interactions could have been guided by this kind of altruism, which compelled all community members to share what they hunted or gathered. If any member accumulated more than others, it would result in a gain at the expense of the rest of the group—in other words, it would be a zero-sum game. If this were the case, selfish behavior might have warranted social sanctions; for example, tribe members might refuse to cooperate in the future with those who behaved selfishly, thereby endangering their own survival. Morality based on reciprocal altruism is essentially intentional; it requires community members to know the intentions of their comrades to identify whether they are willing to cooperate. Consequently, reciprocal altruism can only function in small societies, and it is not surprising that it persists today in small social groups, such as families.

However, economic interactions in our contemporary society are very different because they are based on trade and industrial production, which create conditions conducive to positive-sum games. As we highlighted, the coordination of economic activities in large modern societies involves interactions among thousands or millions of individuals who do not know each other. Under these circumstances, formal legal systems and impersonal institutional frameworks have been developed to regulate the establishment of contracts and ensure the legal security of economic agents. This is why the reciprocal altruism of primitive tribes simply does not scale

up—it is impossible to track the intentions of thousands or millions of individual agents participating in these vast systemic processes. Characterizing modern societies as "one big family" may be a rhetorically appealing metaphor, but it is entirely inapplicable in descriptive and explanatory terms.

5. POLITICAL IMPLICATIONS

Regardless of whether folk beliefs and the moral intuitions of laypeople in economics have evolutionary roots, the issue now is to see what repercussions they may have in the political sphere. As I pointed out at the beginning of this text, beliefs not only provide subjects with descriptions of the world to help them cope with uncertainty, but they also establish a powerful motivation for action. Actions that affect politics and society as a whole are of greater relevance because they impact the lives of millions of people. The impact that beliefs and knowledge have on politics has become one of the most interesting topics for epistemology and cognitive sciences (Haidt 2012; Hannon & de Ridder 2021).

However, not all folk beliefs seem to have the same social impact. Folk linguistic beliefs, such as those described by Jackendoff, may lead speakers to form erroneous views about language, but they do not appear to have significant repercussions. In other cases, the negative effects of some beliefs are more evident: lay conceptions in medicine can lead certain individuals to take inappropriate treatments that may endanger their health or even their lives. Precisely, the emphasis of scientists on promoting scientific knowledge and publicly refuting erroneous beliefs is related to the goal of informing people to prevent them from falling victim to hoaxes or frauds that can affect their finances or even their health. Thus, the negative impact that folk beliefs can produce is variable, as it seems to depend on the specific issues at stake (such as people's mental or physical health).

But the case of folk economic beliefs is different because they directly affect public policies. For example, the belief that the economy is a zero-sum game has produced certain xenophobic stances, such as the claim that foreign immigrants

"steal" local jobs. This can lead to the establishment of discriminatory policies and draconian legislation that would harm not only immigrants but the economy as a whole. This and other examples are clear indicators that lay economic conceptions can impose terrible costs on society.

Economists have been aware of this problem, but they have somewhat dismissed it by appealing to what they call rational ignorance. Based on the "epistemic comparative advantage" described earlier, many economists argue that it is most rational to specialize in the theoretical or practical knowledge that is most relevant to an individual's profession. We exemplified this with the case of the surgeon, who would certainly benefit from dedicating his time and energy to deepening his knowledge and skills in his clinical specialty. As a result, the surgeon may lack other knowledge that has significant social relevance, such as politics or economics.

In this regard, Caplan (2007) points out that rational ignorance is not problematic in the private sphere, but it can be in the public sphere. Returning to our example, the surgeon might adopt an anti-immigration xenophobic stance and, if fully convinced of this belief, might vote for candidates who promote such discriminatory policies. As Caplan argues, rational ignorance entails certain dangers because voters may elect leaders whose platforms or political programs are highly detrimental.

The crucial point is to understand that the primary incentive for politicians in a democracy is to obtain the highest number of votes. Aristotle and the rhetoricians of antiquity already warned us that arguments based on pathos and ethos are the most persuasive in the public arena. To gain empathy with the electorate, the politician will not seek to change the audience's beliefs; instead, he will adapt his discourse according to the audience's demand, which involves incorporating the most widespread beliefs among voters (and if he adorns this discourse with tropes appealing to the public's *ethos* and *pathos*, persuasive success is likely guaranteed). Similarly, if certain folk economic beliefs are widespread among the electorate, it is unsurprising that politicians adopt them, motivating them to promise measures consistent with such beliefs. It matters little whether the politician truly believes in what he proposes; what matters is that folk beliefs create incentive structures that encourage politicians to pass certain laws or government programs, which may have

adverse economic consequences. This occurs because the politician bears no cost for being wrong, as the negative effects of the public policies he has promoted may appear in the medium or long term, and by then, the politician will have moved to another position in the government's hierarchy.

This may be one of the factors explaining the recent surge of authoritarian populism in various regions of the world. This phenomenon has alarmed political science experts due to the demagogic discourse and the types of public policies defended by populist governments. However, what these experts often overlook is that the populist demagogue has merely adapted to the people's beliefs, and if he enjoys broad popularity, it is precisely because he has managed to win their sympathy very effectively. If anything, this phenomenon of authoritarian populism shows that the voters themselves can undermine democracy.

After all, the lay public that adopts these folk economic beliefs is not merely a passive spectator of economic processes but an active economic agent. As such, they can suffer the consequences of economic recessions, such as losing their business or job. The problem arises when they attribute unemployment to foreign immigrants or adopt "conspiracy theories" about the origins of social ills, which can lead them to support certain leaders who implement harmful policies for the entire economy. For example, the populist attempt to revive CEPAL-style mercantilist policies represents a significant danger in our Latin American context. Therefore, the importance of folk economic beliefs is substantial, given their implications for democracy and public policies.

6. CONCLUSIONS

Ultimately, studying lay conceptions helps clarify how humans perceive the world and social reality. In particular, thoroughly investigating folk economic beliefs could shed light on how people without economic training perceive the dynamics of markets and productive activities. This could be revealing about how they react to certain circumstances generated by economic performance, such as economic cycles.

More generally, it can provide important insights for understanding processes of epistemic justification, which has been one of the central problems of contemporary analytic epistemology. These issues can also be of great importance to the cognitive sciences, especially regarding the relationship between cognitive biases and beliefs. In other words, research on folk beliefs could contribute to understanding certain cognitive processes that motivate thought and action.

Another relevant aspect of this research program is that it could explain certain aspects of decision-making processes and political preferences. If beliefs can motivate certain decisions, then studying folk economics would enable the explanation of many of the behaviors of individuals in society. This could be useful for research aiming to investigate the causes behind the political preferences of the electorate. As mentioned in the previous section, the troubling rise of authoritarian populism in Latin America and other regions of the world could be motivated by these kinds of beliefs, among other factors.

In sum, as there is increasing convergence between cognitive sciences, epistemology, and political studies, the topic of lay conceptions could represent a small piece that contributes to the construction of the larger puzzle.

REFERENCES

- Ang, Yuen Yuen. *How China Escaped the Poverty Trap*. Cornell University Press, 2016.
- Ayau Córdón, Manuel F. *El proceso económico. Descripción de los mecanismos espontáneos de la cooperación social*. Editorial Centro de Estudios Económicos Sociales, 2003.
- Baron, Jonathan. *Thinking and Deciding*. Cambridge University Press, 2007.
- Bastiat, Frédéric. “Lo que se ve y lo que no se ve, 1850”. *Obras escogidas*. Madrid: Unión Editorial, 2004. 22-58.
- Bhalla, Surjit S. *The New Wealth of Nations*. Nueva Deli: Simon and Schuster, 2018.

- Bhagwati, Jagdish y Arvind Panagariya. *Why Growth Matters: How Economic Growth in India Reduced Poverty and the Lessons for Other Developing Countries*. Nueva York: Public Affairs, 2013.
- Boyer, Pascal., y Petersen, Michael B. “Folk-economic Beliefs: An Evolutionary Cognitive Model”. *Behavioral and Brain Sciences*. Cambridge University Press, 2018. 1-65. <<https://doi.org/10.1017/s0140525x17001960>>
- Brown, Cecil H. “Folk Botanical Life-forms: Their Universality and Growth”. *American Anthropologist* 79.2 (1977): 317-342. <<https://doi.org/10.1525/aa.1977.79.2.02a00080>>
- _____. “Folk Zoological Life-forms: Their Universality and Growth”. *American Anthropologist* 81.4 (1979): 791-817. <<https://doi.org/10.1525/AA.1977.79.2.02A00080>>
- Buchanan, James M. “Politics Without Romance: A Sketch of Positive Public Choice Theory and Its Normative Implications”. *The Collected Works of James M. Buchanan*. Indianápolis: Liberty Fund, 1999. 45-69.
- Caplan, Bryan. “What Makes People Think Like Economists? Evidence on Economic Cognition from the ‘Survey of Americans and Economists on the Economy’”. *Journal of Law and Economics* 44.2 (2001): 395-426. <<https://doi.org/10.1086/322812>>
- _____. “Sociotropes, Systemic Bias, and Political Failure: Reflexions on the Survey of Americans and Economists on the Economy”. *Social Science Quarterly* 83.2 (2002a): 416-435. <<https://doi.org/10.1111/1540-6237.00092>>
- _____. “Systematic Biased Beliefs About Economics. Robust Evidence on Judgmental Anomalies from the Survey of Americans and Economists on the Economy”. *The Economic Journal* 112.479 (2002b): 433-458. <<https://doi.org/10.1111/1468-0297.00041>>
- _____. “How Do Voters form Positive Economic Beliefs? Evidence from the Survey of Americans and Economists on the Economy”. *Public Choice* 128.1 (2006): 367-381. <<https://doi.org/10.1007/s11127-006-9026-z>>

- _____. The Myth of the Rational Voter. Why Democracies Choose Bad Policies. Princeton University Press, 2007. <<https://doi.org/10.1515/9781400828821>>
- Coleman, William Oliver. Economics and Its Enemies. Two Centuries of Anti-Economics. Nueva York: Palgrave Macmillan, 2002.
- Deaton, Angus. The Great Escape. Health, Wealth, and the Origins of Inequality. Princeton: Princeton University Press, 2013.
- Descartes, René. Meditaciones metafísicas (y otros textos). Tomo II. Madrid: Gredos, 2011.
- Elster, Jon. Una introducción a Karl Marx. Ciudad de México: Siglo XXI, 1991.
- Evans, Jonathan y Keith Frankish. In Two Minds: Dual Processes and Beyond. Oxford University Press, 2012.
- Friedman, Milton y Rose Friedman. Free to Choose: A Personal Statement. Nueva York: HBJ, 1980.
- Gwartney, James D. et ál. Common Sense Economics. What Everyone Should Know About Wealth and Economics. Nueva York: St. Martin's Press, 2016.
- Haidt, Jonathan. The Righteous Mind. Why Good People Are Divided by Politics and Religion. Nueva York: Pantheon Books, 2012.
- Hannon, Michael y Jeroen de Ridder. The Routledge Handbook of Political Epistemology. Londres: Routledge, 2021.
- Hayes, Patrick J. "The Naive Physics Manifesto". Expert Systems on Micro-electronic Age. Ed. Donald Michie. Edinburgh University Press, 1979.
- _____. "The Second Naive Physics Manifesto". Formal Theories of the Commonsense World. Eds. Hobbs, Jerry R. y Moore Robert C., Menlo Park CA: SRI International, 1983. 1-36. <<https://doi.org/10.1016/B978-1-4832-1447-4.50010-9>>
- Hazlitt, Henry. Economics in One Lesson. Nueva York: Harper & Brothers, 1946.
- Heyne, Paul. "Are Economists Basically Immoral?". 'Are Economists Basically Immoral?' and Other Essays on Economics, Ethics, and Religion. Eds. Brennan, Geoffrey y Waterman, A.M.C. Indianápolis: Liberty Fund, 2008. 1-9.
- Hutt, William H. Politically Impossible...? An Essay on the Supposed Electoral Obstacles Impeding the Economical Analysis into Politics or Why Politicians Do

- Not Take Economic Advice. Westerham Kent: Institute of Economic Affairs, 1971.
- Jackendoff, Ray. "The Structure of Language. Why it Matters to Education". Conferencia en Learning and Brain. Cambridge, 5-8 noviembre 2003. <<https://peeps.unet.brandeis.edu/~jackendo/StructureofLanguage1.pdf>>
- Jevons, William Stanley. "Brief Account of a General Mathematical Theory of Political Economy". Journal of the Royal Statistical Society, vol. XXIX, 1862. 282-287.
- Kahneman, Daniel et ál. Judgement under Uncertainty: Heuristics and Biases. Cambridge University Press, 1982. <<https://doi.org/10.1126/science.185.4157.1124>>
- Kahneman, Daniel. et ál. Heuristics and Biases. The Psychology of Intuitive Judgement. Cambridge University Press, 2002.
- Lal, Deepak. Poverty and Progress. Realities and Myths about Global Poverty. Washington: Cato Institute, 2013.
- Mankiw, Gregory N. Principles of Economics. Boston: Cengage, 2021.
- Menger, Carl. Grundsätze der Volkswirthaftslehre. Viena: Braumüller, 1871.
- Newcomb, Simon. "The Problem of Economic Education". The Quaterly Journal of Economics 7.4 (1893): 375-399. <<https://doi.org/10.2307/1882282>>
- Niemietz, Kristian. Socialism: The Failed Idea That Never Dies. Londres: IEA, 2019.
- Parkin, Michael. Economics. Essex: Pearson, 2015.
- Peirce, Charles S. "The Fixation of Belief, 1877". Peirce on Signs. Ed. James Hoopes. University of North Carolina, 1991. 144-159.
- Pinker, Steven. The Blank Slate. The Modern Denial of Human Nature. Nueva York: Penguin Books, 2002.
- _____. Enlightenment Now. The Case for Reason, Science, Humanism, and Progress. Nueva York: Penguin Books, 2018.
- Preston, Carol. "Folk Linguistics". Encyclopedia of Language and Linguistics 2ª ed. Eds. Brown, Keith. Ámsterdam: Elsevier, 2006. 521-533.
- Price, Henry H. "Some Considerations About Belief". Proceedings of the Aristotelian Society 35: Oxford: Oxford University Press, 1934 – 1935. 229-252.

- Rothbard, M. N. "Ludwig von Mises, su esencia". Ludwig von Mises. Infatigable luchador contra la economía ficticia. Ciudad de México: Centro de Estudios en Economía y Educación, A.C. 1983. 20-52.
- Rubin, Paul H. "Folk Economics". *Southern Economic Journal* 7.1 (2003): 157-171. <<https://doi.org/10.1002/j.2325-8012.2003.tb00561.x>>
- Schettino, Macario. *Introducción a la economía para no economistas*. Pearson Educación de México, 2002.
- _____. *Economía en un día*. Ciudad de México: Paidós, 2015.
- Smith, Adam. *An Inquiry into the Nature and Causes of the Wealth of the Nations*, 1776. Indianápolis: Liberty Fund, 1976.
- Smith, Barry y Roberto Casati. "Naive Physics". *Philosophical Psychology* 7.2 (1994): 227-247. <<https://doi.org/10.1080/09515089408573121>>
- Sowell, Thomas. *Economic Facts and Fallacies*. Nueva York: Perseus Books Group, 2011.
- _____. *Basic Economics*. Nueva York: Perseus Books Group, 2015.
- Von Böhm-Bawerk, Eugen. *Kapital und Kapitalzins*. Innsbruck: Wagner, 1884.
- Walras, Léon. *Éléments d'économie politique pure*. Lausana: Corbaz, 1874.
- Wheelan, Charles. *Naked Economics. Undressing the Dismal Science*. Nueva York: W.W. Norton & Company, 2019.
- Wolpert, Lewis. *The Unnatural Nature of Science*. Cambridge, MA: Harvard University Press, 1992.

ELEMENTOS PARA UNA REVISIÓN EPISTEMOLÓGICA DE LA AXIOMÁTICA DE KOLMOGOROV*

ELEMENTS FOR AN EPISTEMOLOGICAL REVISION OF KOLMOGOROV'S AXIOMATICS

ALBERTO LANDRO**

Centro de Investigaciones en Econometría -
Universidad de Buenos Aires y Departamento de
Investigaciones - Universidad del CEMA
Ciudad de Buenos Aires, Argentina.
alandroar@yahoo.com.ar
<https://orcid.org/0009-0000-4402-5726>

MIRTA L. GONZALEZ

Centro de Investigaciones en Econometría -
Universidad de Buenos Aires; Universidad del
CEMA y Gerencia de Estudios Económicos del
Banco Central de la República Argentina
Ciudad de Buenos Aires, Argentina.
mirtagonzalezar@yahoo.com.ar
<https://orcid.org/0009-0001-1186-6104>



RESUMEN

La literatura sobre la probabilidad ha considerado tradicionalmente ciertos supuestos acerca de cuáles fueron los objetivos y alcances de los Grundbegriffe der Wahrscheinlichkeitsrechnung en el desarrollo de la teoría de la probabilidad. Sin embargo, un análisis detallado de la obra de Kolmogorov, en particular en lo referido a su transcurso por las interpretaciones frecuentista y propensionalista, permitió concluir que las hipótesis planteadas habitualmente en la literatura acerca de que su axiomática se basa en un concepto primitivo de la probabilidad, independiente de cualquier interpretación acerca de la na-

* Este artículo se debe citar: Landro, Alberto & Gonzalez Mirta L. "Elementos para una revisión epistemológica de la axiomática de Kolmogorov". *Revista Colombiana de Filosofía de la Ciencia* 24.48 (2024): 79-107. <https://doi.org/10.18270/rcfc.4412>

** Las opiniones expresadas en este artículo pertenecen a los autores y no reflejan los puntos de vista de las instituciones.

turalidad del azar, son en general inexactas y que estos desaciertos en la interpretación condujeron a planteos conceptuales y epistemológicos inadecuados.

Palabras clave: epistemología de la probabilidad; teoría de la probabilidad; interpretación propensionalista-frecuencista del concepto de probabilidad; axiomática de Kolmogorov; teoría de la medida; teoría de grupos.

ABSTRACT

Although the literature on probability has traditionally considered certain assumptions about what were the aims and scope of the "Grundbegriffe der Wahrscheinlichkeitsrechnung" in the development of probability theory, a detailed analysis of Kolmogorov's work in particular regarding its passage through the frequentist and propensionalist interpretations, leads to the conclusion that hypotheses usually raised in the literature about its axiomatics being based on a primitive concept of probability, independent of any interpretation about the nature of chance, are generally inaccurate and that these errors in interpretation led to inadequate conceptual and epistemological approaches.

Keywords: epistemology of probability; probability theory; propensionalist-frequentist interpretation of the concept of probability; Kolmogorov's axiomatics; measure theory; group theory.

1. *DIE GRUNDBEGRIFFE DER WAHRSCHEINLICHKEITSRECHNUNG*

Según la literatura sobre la probabilidad, los objetivos de la propuesta de Andrei Nikolaievitch Kolmogorov —contenidos en los *Grundbegriffe* (1933)— consistían en 1) solucionar el problema de la falta de rigurosidad con que los matemáticos de los siglos XVIII, XIX y la primera mitad del XX habían tratado los problemas de la probabilidad; 2) considerar la probabilidad una idea primitiva, independiente de cualquier interpretación acerca de la naturaleza del azar y 3) a partir de la concepción de la teoría de la probabilidad como un caso particular de una teoría matemática abstracta, definir una axiomática que actuara como sistema de referencia de las axiomáticas inherentes a las distintas interpretaciones sobre la naturaleza del azar.

El análisis cuidadoso de la evolución del pensamiento probabilístico de Kolmogorov realizado en este trabajo permitió concluir que algunas de estas apreciaciones explícitas o implícitas de la literatura utilizada habitualmente en los cursos de pregrado sobre los alcances teórico-conceptuales de los *Grundbegriffe* son, en general, opinables y que sus interpretaciones sobre los fundamentos filosóficos de la axiomática afectaron los alcances de los teoremas derivados.

2. UN ANÁLISIS DE LOS ANTECEDENTES

2.1. LA TEORÍA DE LA MEDIDA

Los primeros intentos de vinculación de la teoría de la medida de conjuntos y la integración de Lebesgue con la cuantificación de la probabilidad fueron propuestos por Borel, iniciando el proceso que culminó con el sistema axiomático contenido en los *Grundbegriffe*.

Mediante la formulación de la conocida como ley fuerte de los grandes números, Borel (1909a, 1909b) logró formalizar la relación entre la teoría de la medida y una interpretación geométrica que constituye el núcleo de la teoría clásica de la probabilidad. Pero su justificación teórica de esta relación, utilizando generalizaciones a conjuntos numerables de los teoremas de la probabilidad total y de la probabilidad compuesta, resulta innegablemente incompleta debido a circunstancias atribuibles posiblemente a razones filosóficas relacionadas con su rechazo a aceptar la noción de probabilidad numerable, las cuales lo obligaron a emplear argumentos frecuentistas que consideran a las probabilidades límites de sucesiones de repeticiones independientes.

Fueron Faber (1910) y Hausdorff (1914) quienes enmendaron esta limitación, de lo cual obtuvieron demostraciones rigurosas de la ley fuerte de los grandes números utilizando exclusivamente argumentos de teoría de la medida, generalizados —como se verá más adelante— por Cantelli (1916a, 1916b, 1917a)¹. El resultado de estas propuestas fue una ley fuerte de los grandes números interpretable según dos versiones: una relacionada con el ámbito de los números reales y otra con la probabilidad.

¹ Los resultados obtenidos por Cantelli fueron de gran importancia en la medida en que impulsaron a otros autores al desarrollo de diferentes definiciones de convergencia estocástica (véase Doob 1989, 1996; Von Plato 1994).

Siguiendo los cánones de la axiomatización de Sierpiński (1918) y de la medida de Lebesgue (1901, 1904)², Steinhaus (1923, 1930a, 1930b) en sus trabajos sobre los teoremas en el límite propuso una solución a esta dualidad conceptual mediante la axiomatización de la teoría de Borel³:

Sean A el conjunto de todas las sucesiones binarias infinitas de eventos (es decir, que admiten solo dos resultados posibles), C una clase de subconjuntos de A y una función real p que representa las probabilidades de los elementos de un subconjunto E de C , *axioma 1*: $p(E) \geq 0$; *axioma 2*: a) para toda sucesión binaria finita E^* , el subconjunto E de A formado por todas las sucesiones binarias infinitas que comienzan con E^* pertenecen a C ; b) si dos de tales sucesiones binarias finitas, E_1^* y E_2^* , difieren solamente en un elemento entonces $p(E_1) = p(E_2)$ (donde E_1 y E_2 denotan los subconjuntos de A que comienzan con las sucesiones E_1^* y E_2^* , respectivamente); c) $p(A) = 1$; *axioma 3*: la clase C es cerrada respecto a uniones finitas y numerables de elementos disjuntos, y la función p satisface las condiciones de aditividad finita y aditividad numerable; *axioma 4*: si $E_1 \supset E_2$ y $E_1, E_2 \in C$, entonces, $(E_2 / E_1) \in C$; *axioma 5*: si $E \in C$ y $p(E) = 0$, entonces para todo subconjunto E^* tal que $E \supset E^*$, $p(E^*) = 0$.

La axiomática de Steinhaus incluye los axiomas 1, 3, 4 y 5 de Sierpiński y un axioma (que Steinhaus demostró que es equivalente a su axioma 2), según

² En su expresión axiomática de la teoría de la medida, Sierpiński (1918) caracterizó la clase de conjuntos medibles de Lebesgue como la menor clase C^* de conjuntos que satisfacen las siguientes condiciones: *axioma 1*: a cada conjunto $E \in C$ es posible asignarle un número real $p(E)$ que defina su medida y satisfaga los axiomas que siguen; *axioma 2*: todo intervalo finito y cerrado $E \in C^*$ posee una longitud y una medida; *axioma 3*: la clase C^* es cerrada para cualquier unión finita y numerable de uniones de elementos disjuntos y p es una medida finita y aditiva numerable; *axioma 4*: sean dos conjuntos E_1 y E_2 tales que $E_1 \supset E_2$ y que $E_1, E_2 \in C^*$, entonces, $E_1 / E_2 \in C^*$; *axioma 5*: si el conjunto $E \in C^*$ y $p(E) = 0$, entonces, todo subconjunto de $E^* \in C^*$.

³ La memoria de Steinhaus de 1923 dio origen a la conocida como escuela polaca en el estudio de funciones independientes, a la cual se deben importantes aportes a la teoría de la probabilidad.

el cual la probabilidad de que el evento E se encuentre en un intervalo J ($p(E \subset J)$) es igual a la medida de dicho intervalo⁴.

Posteriormente Khintchine (1929, 1932a, 1932b), Kolmogorov (1929, 1931, 1933) y Prokhorov (1956), entre otros, iniciaron la sucesión de generalizaciones de las condiciones necesarias y suficientes para poder asegurar la validez de la ley de los grandes números⁵.

En particular Kolmogorov (1933) estableció que la diferencia entre la probabilidad y la teoría de la medida es la condición de independencia. Este postulado fundamental puede ser considerado el punto de partida “[...] hacia un consenso significativo acerca de los aspectos matemáticos de la probabilidad [...] en tanto que su interpretación filosófica transcurre hacia la alternativa de la probabilidad de ser epistémica u objetiva” (De Schewemaerkere & Szafarz 2008 2)⁶.

2.2. LA TEORÍA DE GRUPOS

Otro antecedente importante en el proceso de construcción de los *Grundbegriffe* y, por lo tanto, de la formulación de la axiomática de Kolmogorov fue el trabajo de Slutsky (1922): “Slutsky fue el primero en dar una versión correcta del contenido puramente matemático de la teoría de la probabilidad” (Kolmogorov 1948 143). Su propuesta asumió como punto de partida la sustitución del término *probabilidad* por el de *valencia*⁷. En su concepto, la probabilidad sería solo una representación del cálculo de valencias. Sustituyó el concepto de equiprobabilidad por el simple supuesto de que si determinados casos a los cuales se les

⁴ Véase Holgate (1997).

⁵ Véase Landro y González (2016).

⁶ El desarrollo de las teorías de la medida y de la integración de Lebesgue a finales del siglo XIX y comienzos del XX ya ha sido estudiado exhaustivamente (véase, por ejemplo, Hawkins, 1970; Pier, 1994a, 1994b). Por lo tanto, aquí se intentará solamente una breve reseña de aquellos trabajos que influyeron directamente en el contenido de los *Grundbegriffe*.

⁷ Laemmel (1904) ya había utilizado el término alemán *valenz*.

asigna una valencia α son subdivididos, las valencias asignadas a los subcasos deberían sumar α (pudiendo ser las valencias asignadas a los subcasos iguales o no). De esta forma, los teoremas de la probabilidad total y de la probabilidad compuesta pasaron a pertenecer a un cálculo (que conceptualmente no puede ser considerado de probabilidades) tan abstracto como la teoría de grupos.

En su memoria de 1929, Kolmogorov adoptó el planteo de Slutsky (1925) consistente en concebir la probabilidad como un caso particular de una teoría abstracta y propuso una interpretación de la probabilidad relacionada con una función $M(\cdot)$ (a la que denominó *especificación de la medida*) que asigna números no-negativos $M(E)$ a cada elemento E de una clase de subconjuntos de un conjunto de atributos A ; además, supuso solamente que dados dos elementos E_1 y E_2 disjuntos, entonces la función $M(\cdot)$ asignará un número a dos cualesquiera de los conjuntos E_1 y E_2 , de modo que el número a asignar al tercer conjunto $E_1 \cup E_2$ queda determinado por añadidura, verificándose que⁸

$$M(E_1 \cup E_2) = M(E_1) + M(E_2).$$

Con respecto a la probabilidad de ocurrencia de E_2 condicionada por la supuesta ocurrencia de E_1 , Kolmogorov propuso la siguiente definición, basada en los resultados obtenidos por Steinhaus:

$$M(E_2 / E_1) = \frac{M(E_1 \cap E_2)}{M(E_1)}.$$

En 1930, como una extrapolación lógica al caso de sucesiones infinitas de eventos, Kolmogorov demostró que las únicas probabilidades que satisfacen los axiomas de Steinhaus son las que se obtienen de un sistema isomórfico con la medida

⁸ Obsérvese que estas condiciones coinciden con los axiomas 3 y 4 de Steinhaus (1923).

de Lebesgue en el intervalo $[0, 1]$ y, en consecuencia, con la axiomática de Sierpiński. Su extensión del alcance de estos axiomas incluyó todas las sucesiones de variables aleatorias, aun las variables complejas, pero no trascendió el contexto geométrico y, por lo tanto, no logró formular una noción de probabilidad en espacios abstractos. Esta tarea les correspondió a los matemáticos Stanislaw Ulam y Zbigniew Łomnicki, quienes demostraron en sus memorias de 1923, 1932 y 1934 que $M(\cdot)$ es una medida de probabilidad sobre una σ -álgebra completa, y que, a partir de una sucesión numerable de espacios, con medidas de probabilidad M , se puede construir una medida de probabilidad que satisfaga las mismas condiciones en el espacio definido por el producto cartesiano de los espacios de los eventos marginales.

El hecho de que las operaciones entre eventos puedan ser identificadas con las operaciones entre conjuntos de tal modo que subsistan las propiedades formales de estas últimas permite quitar a los eventos toda base material y considerarlos entes matemáticos susceptibles de aplicarles procedimientos, razonamientos y deducciones propios de la lógica matemática. Sin embargo, más allá de su innegable importancia, estos resultados sobre los avances de la justificación de la teoría de la probabilidad permiten concluir que si bien la representación de los eventos mediante conjuntos constituye un arbitrio útil para facilitar su tratamiento matemático, no permite ninguna interpretación acerca de la naturaleza de estos. En otras palabras, una estructura matemática abstracta no puede aprehender el contenido intuitivo que posee el concepto de evento.

2.3. LA ABSTRACCIÓN FRECUENCISTA

Por su parte, Cantelli (1917b, 1932) introdujo una teoría abstracta de la probabilidad en la que descarta el empleo de nociones vinculadas a cuestiones empíricas como las de *posibilidad*, *evento*, *probabilidad* e *independencia*, y en la que solo formalmente conserva las reglas clásicas de la probabilidad total y compuesta. Denotando por $m(E)$ la medida de un conjunto E , estableció que: i) dados dos subconjuntos E_1 y E_2 disjuntos, se verifica que

$$M(E_1 \cup E_2) = M(E_1) + M(E_2);$$

ii)

$$0 \leq \frac{m(E_1 \cap E_2)}{m(E_i)} \leq 1 \quad (i = 1, 2),$$

en particular si se verifica que $m(E_1 \cap E_2) = m(E_1)m(E_2)$, considera que los conjuntos E_1 y E_2 son “*multiplicables*”; y iii) aún la ley de los grandes números de Bernoulli y sus generalizaciones y la ley del logaritmo iterado de Khinchin⁹ pueden ser tratadas con el mismo nivel de abstracción que propone su teoría.

Como respuesta a la consideración de Slutsky según la cual, dado que las sucesiones empíricas de repeticiones independientes poseen propiedades que

⁹ Como una generalización de la ley de los grandes números Kintchine (1924) demostró la ley del logaritmo iterado, según la cual, dada una sucesión de variables aleatorias $Y_i: b(l, \theta)$, existe una función: $f(n) = \sqrt{2n\theta(1-\theta)\log[\log(n)]}$ que caracteriza la tasa de convergencia, tal que, para todo ε, δ , un valor n_0 y una variable $Y_{(n)} = \sum_{i=1}^n Y_i$, se verifica que $p\left(\left|\frac{Y_{(n)} - \frac{n}{2}}{f(n)}\right| < 1 + \varepsilon\right) = 1 - \delta$ (ver Landro; González (2016)).

trascienden la existencia de un límite, la probabilidad no podía ser asimilada al límite de una frecuencia relativa, Cantelli (1916a, 1916b, 1917a, 1935) desarrolló, en forma independiente, un proyecto de justificación de la interpretación frecuencista de la probabilidad según su sistema axiomático abstracto. Según su planteo, el cálculo de probabilidades clásico —que considera la clase particular de eventos que se caracterizan porque el mecanismo físico que genera la aleatoriedad incluye en forma simétrica a todos los resultados posibles— debía ser desarrollado (en forma circular) en tres etapas: la primera consistía en comprobar experimentalmente el supuesto de equiprobabilidad (comprobando si ocurren con la misma frecuencia) y de este modo justificar empíricamente el empleo de las reglas de la probabilidad total y compuesta; la segunda etapa estaba dirigida a desarrollar, sin hacer referencia a su justificación empírica, una teoría abstracta basada exclusivamente en su axiomática; y la tercera consistía en deducir las probabilidades de dicha teoría y utilizarlas para predecir las frecuencias relativas.

3. LA AXIOMÁTICA

Como una culminación del desarrollo experimentado por la teoría de la medida entre 1900 y 1930, se publicó *Grundbegriffe der Wahrscheinlichkeitsrechnung*:

El momento en el que se resumieron todos los elementos necesarios para formular explícitamente el cuerpo completo de axiomas de la teoría de la probabilidad (clásica modernizada) fue, precisamente, aquél en que el señor Borel introdujo esta nueva clase de aditividad [numerable] en el cálculo de probabilidades —en 1909—. No basta con poseer todas las ideas ‘in mente’, se debe estar seguro que esa totalidad de ideas es sufi-

ciente, expresarlas en forma conjunta y asumir la responsabilidad de afirmar que no es necesario nada más para construir la teoría. Esto fue lo que hizo el señor Kolmogorov. Este fue su logro. (Fréchet 1930a 1059)

En sus 62 páginas Kolmogorov logró introducir toda la matemática sintetizada fundamentalmente en la obra de Borel, Steinhaus, Slutsky, Cantelli y Daniell (en particular en lo relacionado con su teorema de consistencia, es decir, con la construcción de probabilidades en espacios de infinitas dimensiones), Radon (1913) y Nikodym (1930) (en lo referido a las probabilidades condicionadas y a los valores esperados), y, en particular, en la teoría de Fréchet (1930b) (respecto a la condición de independencia estocástica)¹⁰.

En respuesta al sexto de la famosa colección de problemas propuestos por Hilbert según el cual “Debemos tratar mediante axiomas aquellas ciencias físicas en las cuales las matemáticas juegan un rol importante, en particular, la teoría de la probabilidad y la mecánica” (1902 237), Kolmogorov adoptó una posición filosófica muy semejante a la justificación utilizada por Fréchet con respecto a la necesidad de independizar el concepto de probabilidad de la interpretación geométrica de Borel, Lévy y Slutsky.

Esta transición de la interpretación geométrica a la propuesta por Fréchet fue considerada por Kolmogorov una base fundamental de la teoría de la probabilidad y constituyó el argumento para la formulación de lo que, arbitrariamente, muchos probabilistas denominan la axiomática clásica de la probabilidad¹¹:

¹⁰ Cabe mencionar que en el núcleo de la bibliografía de los *Grundbegriffe* figuran autores como Sergei Bernstein, Emile Borel, Paul Lévy, Antoni Lomnicki, Evgeny Slutsky, Hugo Steinhaus, Richard von Mises, Guido Castelnuovo y Maurice Fréchet. A esta lista cabría agregar los aportes de Henri Lebesgue, Waclaw Sierpiński, Constantin Carathéodory, Johann Radon, Otton Nikodym, Percy Daniell y Norbert Wiener, cuya obra constituye la herencia filosófica y matemática que recibió Kolmogorov.

¹¹ Esta consideración se evidencia en la apología que Kolmogorov realiza en el prefacio de los *Grundbegriffe* de los resultados obtenidos por Fréchet: “Después de las investigaciones de Lebesgue, las analogías entre la medida de un conjunto y la probabilidad de un evento y la integral de una función y la

- *Axioma 1:* los eventos forman una σ -álgebra s , es decir, una clase cerrada respecto a las operaciones de unión, intersección y negación de conjuntos numerables de eventos y del límite de sucesiones de eventos, es decir: a) si los eventos $E_j \in s$ ($j = 1, 2, \dots$), entonces, $\bigcup_{j=1}^{\infty} E_j \in s$; y b) si $E_j \in s$ ($j = 1, 2, \dots$), entonces, $\bigcap_{j=1}^{\infty} E_j \in s$ (en realidad, dado que $E_i \cap E_j = (E_i \cup E_j) - [(E_i - E_j) \cup (E_j - E_i)]$, esta propiedad puede considerarse una consecuencia del postulado a); y c) dada una sucesión de eventos $\{E_1, E_2, \dots\}$, pertenecientes a s , entonces $\lim_{j \rightarrow \infty} E_j \in s$.
- *Axioma 2:* dado un espacio muestral Ω , se verifica que $\Omega \in s$.
- *Axioma 3:* asociado a cada evento $E \in s$, existe un número real no-negativo, $p(E)$, al que se denominará *probabilidad de ocurrencia del evento E*.
- *Axioma 4:* la probabilidad de que ocurra al menos uno de los eventos incluidos en el espacio muestral es igual a 1: $p(\Omega) = 1$.
- *Axioma 5 (de aditividad):* sean E_1 y E_2 dos eventos incompatibles, es decir, tales que no pueden presentarse en forma simultánea ($E_1 \cap E_2 = \emptyset$), entonces se verificará que

$$p(E_1 \cup E_2) = p(E_1) + p(E_2)$$
 (la primera formulación estricta de este principio figura en el *Ars conjectandi* (1713) de J. Bernoulli).
- *Axioma 6 (de continuidad):* dada una sucesión monótona de eventos, $E_i \supset E_{i+1}$ ($i = 1, 2, 3, \dots$), se puede escribir:

$$E_i = E_1 \cup (E_1 \cap \bar{E}_2) \cup \dots \cup (E_{i-1} \cap \bar{E}_i) \quad (i = 2, 3, \dots),$$

esperanza matemática de una variable aleatoria quedaron claras. Estas analogías podrían ser extendidas dado que, por ejemplo, muchas propiedades de las variables aleatorias independientes son análogas a las correspondientes propiedades de las funciones ortogonales. No obstante, para basar la teoría de la probabilidad sobre estas analogías era necesario independizar a la teoría de la medida y a la integración de los elementos geométricos que aún figuraban en primer plano en la propuesta de Lebesgue. Este logro fue obtenido por Fréchet". (1992 p.v)

de lo que resulta que

$$p(E_i) = p(E_1) + \sum_{j=2}^i p(E_{j-1} \cap \bar{E}_j) \quad (j = 2, 3, \dots),$$

de modo que, aplicando miembro a miembro el operador límite, será

$$\begin{aligned} \lim_{i \rightarrow \infty} p(E_i) &= p(E_1) + \sum_{j=2}^{\infty} p(E_{j-1} \cap \bar{E}_j) = p \left[E_1 \bigcup_{j=2}^{\infty} (E_{j-1} \cap \bar{E}_j) \right] = \\ &= p \left\{ \lim_{i \rightarrow \infty} [E_1 \cup_{j=2}^i (E_{j-1} \cap \bar{E}_j)] \right\} = p \left(\lim_{i \rightarrow \infty} E_i \right). \end{aligned}$$

Así, se puede concluir que la probabilidad es una función continua respecto a cualquier sucesión monótona de eventos.

Así mismo, si dicha sucesión es tal que $E_i \supseteq E_{i+1}$ ($i = 1, 2, \dots$), entonces, se verificará que $\bar{E}_i \subseteq \bar{E}_{i+1}$ y, por lo tanto, dada la existencia del límite de la sucesión de las $p(\bar{E}_i)$, será

$$\lim_{i \rightarrow \infty} p(\bar{E}_i) = 1 - \lim_{i \rightarrow \infty} p(E_i) = 1 - p(\bigcup_{i=1}^{\infty} E_i) = p(\bigcap_{i=1}^{\infty} \bar{E}_i) = p \left(\lim_{i \rightarrow \infty} \bar{E}_i \right).$$

Luego, dados los cinco primeros axiomas, se demuestra fácilmente que este sexto axioma de continuidad es equivalente a la condición de *aditividad completa* o *aditividad numerable* o σ -*aditividad* (que, obviamente, contiene a la aditividad simple como caso particular): sea $\{E_1, E_2, \dots, E_n, \dots\}$ un conjunto de eventos incompatibles a pares, es decir, tales que $E_i \cap E_j = \emptyset$ ($i \neq j; i, j = 1, 2, \dots$). Por inducción, se demuestra que

$$\bigcup_{j=1}^{n+1} E_j = \bigcup_{j=1}^n E_j + E_{n+1}.$$

Dado que dos cualquiera de los eventos E_j ($j = 1, 2, \dots, n$) y E_{n+1} son incompatibles, será

$$p(\cup_{j=1}^{n+1} E_j) = p(\cup_{j=1}^n E_j) + p(E_{n+1}) = \sum_{j=1}^{n+1} p(E_j).$$

De la misma forma, se puede escribir:

$$\cup_{j=1}^{\infty} E_j = (\cup_{j=1}^n E_j) \cup (\cup_{j=n+1}^{\infty} E_j),$$

y, como cada uno de los eventos E_j ($j = 1, 2, \dots, n$) es incompatible con cada uno de los eventos E_j ($j = n + 1, n + 2, \dots$), se verificará que

$$\begin{aligned} p\left(\bigcup_{j=1}^{\infty} E_j\right) &= p\left(\bigcup_{j=1}^n E_j\right) p\left(\bigcup_{j=n+1}^{\infty} E_j\right) = \\ &= \sum_{j=1}^n p(E_j) + p\left(\bigcup_{j=n+1}^{\infty} E_j\right) = \\ &= \lim_{n \rightarrow \infty} \sum_{j=1}^n p(E_j) + \lim_{n \rightarrow \infty} p(\cup_{j=n+1}^{\infty} E_j) = \sum_{j=1}^{\infty} p(E_j). \end{aligned}$$

Debe tenerse en cuenta que los eventos $(\cup_{n+1}^{\infty} E_j)$ forman una sucesión decreciente, es decir que $(\cup_{n+1}^{\infty} E_j) \subset (\cup_n^{\infty} E_j)$, de modo que $\lim_{n \rightarrow \infty} (\cup_{n+1}^{\infty} E_j) = 0$.

Una lectura analítica permite concluir, entonces, que la axiomática de Kolmogorov puede resumirse en el principio de que $p(\cdot)$ es una función no-negativa, aditiva en el sentido de Fréchet (1937-1938), con el agregado de la condición $p(\Omega) = 1$.

Como corolario de estas consideraciones se puede concluir que los *Grundbegriffe* satisficieron estrictamente el objetivo planteado en su prefacio:

“[...] satisfacer una falencia en la literatura sobre la teoría de la probabilidad (incluyendo en dicha literatura el libro de Fréchet, que se encuentra en preparación) y brindar una presentación del sistema en su totalidad, libre de complicaciones superfluas” (Kolmogorov 1992 p.v).

4. UNA REVISIÓN CRÍTICA DEL CONTENIDO FILOSÓFICO DE LOS *GRUNDBEGRIFFE*

4.1. LA CONDICIÓN DE INDEPENDENCIA

Los *Grundbegriffe* no solo lograron sintetizar el desarrollo experimentado por la teoría de la medida, sino que presentaron avances fundamentales respecto a la matemática de Borel, Fréchet, Daniell y Cantelli: 1) a diferencia del sistema abstracto de Cantelli, Kolmogorov propuso una axiomática asociada al concepto de probabilidad, lo cual permitió desarrollar, a partir de la formulación de un método general de representación de la teoría de la medida, una teoría de la probabilidad concebida como una rama conceptualmente independiente de la matemática; 2) en este sentido, incorpora el concepto de independencia estocástica de los eventos como el factor fundamental de diferenciación entre la probabilidad y la teoría de la medida; y 3) al menos en términos formales, el resultado obtenido por Kolmogorov, en la medida en que considera una cardinalidad arbitraria de los elementos del conjunto, es más general que el proporcionado por Daniell (1921), que considera una cardinalidad numerable —una generalización cuyo carácter exclusivamente conceptual permite concluir que no implica modificaciones matemáticas significativas, ni resulta relevante en las aplicaciones.

4.2. LA TRANSICIÓN FRECUENCISTA

Si bien la axiomática de Kolmogorov posee la capacidad de transformar valores de probabilidad de eventos *simples* en valores de probabilidad de eventos *complejos*, en la medida en que solo *determina* los valores de probabilidad 0 y 1 asignados a los que él define como *eventos imposibles* y *eventos ciertos*, respectivamente, evidencia su incapacidad para asignar valores de probabilidad a eventos no triviales como los eventos *imposibles* o *ciertos*. El intento de solución de esta limitación condujo a Kolmogorov al transcurso hacia una interpretación frecuencista y, posteriormente, a la adopción de una definición acorde con la interpretación propensionalista de Popper.

La circunstancia inequívoca que permite confirmar esta transición a la interpretación frecuencista por parte de Kolmogorov es su aceptación del principio de Cournot y de la axiomática de Von Mises (1931), los cuales postulan la aproximación de los resultados de un conjunto de condiciones repetibles al supuesto verdadero valor de una probabilidad como nexo principal de la axiomática con el mundo real. Esta aproximación está sintetizada en el siguiente párrafo de la sección 2 de los *Grundbegriffe* (de dos páginas de extensión, bajo el título “Das Verhättnis zur Erfahrungswelt”):

Es posible asumir que a un resultado A de ocurrencia eventual bajo ciertas condiciones ζ , que no es nuestra intención analizar aquí, se le puede asignar un número real $p(A)$ que posee las siguientes propiedades: a) puede ser considerado como prácticamente cierto que, si el sistema de condiciones σ se repite un número suficientemente grande de veces (n) y se verifica que el evento A ocurre m veces, entonces el cociente $\frac{m}{n}$ diferirá levemente de $p(A)$; b) si $p(A)$ es muy pequeño, entonces se puede considerar como prácticamente cierto que el evento A no ocurrirá en una repetición individual de las condiciones ζ (1992 4).

El conocido como *principio de Cournot* (1843) está relacionado con la existencia de eventos incluidos en un dominio Ω a los que también se les asigna una probabilidad unitaria: son los denominados eventos *quasi-ciertos* o *ciertos en-probabilidad*. La negación de un evento quasi-cierto define un evento que, no coincidiendo con el evento imposible, se comporta como tal en lo que se refiere a su probabilidad y se denomina *quasi-imposible* o *imposible en-probabilidad*¹². Estos eventos poseen las siguientes propiedades: 1) la unión de dos eventos imposibles-en probabilidad es un evento imposible en probabilidad (propiedad generalizable a cualquier conjunto finito o infinito numerable de eventos); 2) sean E_I un evento imposible en-probabilidad y E un evento ni cierto ni imposible, entonces se verifica que $(E_I \cap E)$ es un evento imposible en probabilidad y que $(E_I \cup E)$ es un evento ni cierto ni imposible; 3) sean E_C un evento cierto en-probabilidad y E un evento ni cierto ni imposible, entonces se verifica que $(E_C \cap E)$ es un evento ni cierto ni imposible y que $(E_C \cup E)$ es un evento cierto en-probabilidad.

Estas definiciones derivan de los conceptos de evento *moralmente imposible* y *moralmente cierto* formulados originalmente por Bernoulli (1713). La *certeza moral probabilística* fue ampliamente discutida en el siglo XVIII. D'Alembert (1761, 1767, 1780) planteó la necesidad de distinguir entre eventos *metafísicamente imposibles* (inmunes a la prueba de la experiencia y *físicamente imposibles*¹³). Al respecto, Buffon (1777) consideró que la diferencia entre las certezas moral y física era una cuestión de grado de aproximación; según su apreciación, un evento con probabilidad $\frac{9999}{10000}$ es *moralmente cierto*, en tanto que un evento con una probabilidad mucho mayor puede ser considerado *físicamente cierto*.

¹² Es decir un evento cuya probabilidad, si bien no es efectivamente igual a cero, solo posee un interés puramente metafísico.

¹³ Véase Daston (1979).

Borel (1909a, 1912, 1925, 1930, 1939), utilizando a menudo un estilo más literario que matemático o filosófico, adoptó el tratamiento de los eventos con probabilidad infinitesimal como imposibles y estableció una clasificación de estos más refinada que la de Buffon, según la cual una probabilidad de 10^{-6} es despreciable *a escala humana*, una probabilidad de 10^{-15} es despreciable *a escala terrestre* y una probabilidad de 10^{-50} es despreciable *a escala cósmica*¹⁴.

Cournot, basándose en el concepto de probabilidad geométrica de Borel, planteó una formulación de la ley de los grandes números según la cual se puede asegurar que un evento moralmente imposible no ocurrirá (que su ocurrencia es matemáticamente posible pero físicamente imposible) y, como corolario, que es matemáticamente imposible que, dada una sucesión suficientemente larga de repeticiones, la frecuencia relativa de un evento moralmente imposible difiera sustancialmente de la probabilidad nula: “[...] el evento físicamente imposible es, por lo tanto, aquél al que le corresponde una probabilidad infinitamente pequeña y esta sola observación es suficiente para proporcionar una sustancia objetiva fenoménica a la teoría matemática de la probabilidad” (158)¹⁵.

La ley de los grandes números de Cournot puede ser expresada según dos modalidades: 1) una *forma fuerte* o *estricta* que considera que un evento con probabilidad infinitesimalmente pequeña no ocurrirá en la repetición siguiente y que, combinada con los postulados del teorema de Bernoulli, permite concluir que la probabilidad de un evento *siempre* puede ser aproximada por su frecuencia relativa a partir de una serie suficientemente larga de repeticiones independientes; y 2) una forma *débil* que considera que un evento con probabilidad muy

¹⁴ A partir de los trabajos de Wiman (1900, 1901) sobre fracciones continuas, los matemáticos alemanes dirigieron sus esfuerzos a demostrar que existen conjuntos de medida nula y que estos conjuntos son *imposibles* en términos probabilísticos, lo cual puso en evidencia el abismo que existe entre la probabilidad nula y la probabilidad infinitesimal (véase Bernstein, 1912).

¹⁵ La expresión *objetiva fenoménica* se refiere a la distinción planteada por Kant entre el *noumenon*, o *cosa en sí misma*, y el *phenomenon*, u *objeto de la experiencia*. Véase Daston (1994).

próxima a cero ocurrirá muy raramente en una serie de repeticiones y que, combinada con el teorema de Bernoulli, permite concluir que la probabilidad de un evento *habitualmente* puede ser aproximada por su frecuencia relativa a partir de una serie suficientemente larga de repeticiones independientes.

Kolmogorov (según los postulados de la *forma fuerte* de la ley de los grandes números), siguiendo a Cournot, asumió que el principio que asimilaba una probabilidad infinitesimalmente pequeña a la imposibilidad de ocurrencia de un evento proporcionaba un significado empírico a la probabilidad clásica, pero no incluyó ninguna consideración acerca de la existencia de la diferencia conceptual entre la afirmación de que un evento de probabilidad infinitesimal no ocurrirá y el supuesto de que, por sí sola, esa afirmación de no-ocurrencia del evento proporciona un significado empírico a la teoría clásica.

4.3. LA INTERPRETACIÓN PROPENSIONALISTA

Los probabilistas franceses y rusos sucesores de Laplace, en desacuerdo con la que se podría denominar la *filosofía ortodoxa* de su interpretación clásica, comenzaron asumiendo una interpretación subjetivista y luego, a partir de una lectura equivocada de la *versión inversa del teorema de Bernoulli*¹⁶, del principio de Cournot y, sin una base conceptual válida del teorema de Bayes, intentaron establecer la existencia de probabilidades objetivas de significado empírico mediante la aplicación de la ley de los grandes números.

A fin de resolver estos errores de interpretación, Von Kries (1886) propuso, como condición necesaria para razonar acerca de probabilidades objetivas de significado empírico a partir de frecuencias relativas, el principio de la equiprobabilidad, que él denominó *principio del Spielraum*, según el cual las probabilidades

¹⁶ Véase Landro, González (2016).

objetivas pueden ser calculadas a partir de casos en los cuales las circunstancias que genera cada caso admiten el mismo número de arreglos posibles.

En lo referido al supuesto de equiprobabilidad como condición necesaria, Hadamard (1922), según una posición netamente determinística, sostuvo que las nociones de eventos igualmente probables (*perfectamente equivalentes*, según su nomenclatura) y de eventos muy improbables constituyen el fundamento de la teoría de la probabilidad, y que el hecho de que la *equivalencia perfecta* sea una condición matemática no verificable no debe impedir la aproximación empírica.

A este respecto, si bien Lévy (1925) coincidió con Hadamard en que la condición de equiprobabilidad puede ser considerada un fundamento de la matemática de las probabilidades, admitió que —dada la imposibilidad empírica de su verificación—, si se la asume como única base de razonamiento, las probabilidades obtenidas serán necesariamente subjetivas.

La revisión de los *Grundbegriffe* permitió concluir que, si bien es innegable que las propuestas de Fréchet y Slutsky ejercieron una influencia significativa sobre el pensamiento de Kolmogorov, debido a las limitaciones del principio de Cournot planteadas por Lévy, Kolmogorov formuló sus axiomas posteriormente justificados por la definición propensionalista de Popper (1957a, 1957b, 1959). Esta consiste en asociar las probabilidades a los resultados de un sistema de condiciones σ repetibles (no especificadas en el texto) y no a los colectivos de la interpretación frecuentista. Como consecuencia de la adopción de esta interpretación popperiana, la vinculación de la axiomática abstracta de Kolmogorov con el mundo de la experiencia se materializa a partir de principios no-operacionalistas, según los cuales los conceptos teóricos fundamentales son asumidos habitualmente como indefinidos y se vinculan con la observación solo en forma indirecta. En “On tables of random numbers” (1963 369), Kolmogorov expresa que la frecuencia límite es una idealización teórica, un concepto de interés puramente matemático:

La noción de probabilidad como límite de la frecuencia relativa no contribuye en nada a la aplicabilidad de los resultados de la teoría de la probabilidad a problemas prácticos reales que sólo admiten un tratamiento a partir de un número finito de repeticiones.

Por otra parte, a partir de los postulados de Borel y Lévy, se concluyó aquí que el ya mencionado principio kolmogoroviano, según el cual

[...] a un resultado A de ocurrencia eventual bajo ciertas condiciones ζ [...] se le puede asignar un número $p(A)$ tal que a) puede ser considerado como prácticamente cierto que, si el sistema de condiciones ζ se repite un número suficientemente grande de veces (n) y se verifica que el evento A ocurre m veces, entonces el cociente $\frac{m}{n}$ diferirá levemente de $p(A)$ (Kolmogorov 1992 4).

puede ser deducido de la consideración conjunta del principio: “b) si $p(A)$ es muy pequeño, entonces se puede considerar como prácticamente cierto que el evento A no ocurrirá en una repetición individual de las condiciones ζ ” (Kolmogorov 1992 4) y del teorema de Bernoulli (el cual surge como un corolario de los axiomas).

Luego, la conclusión de que el principio a) mencionado en el párrafo anterior no solo precede a los axiomas sino que es utilizado para su formulación, y que, por lo tanto, precede al teorema de Bernoulli, permitió, a su vez, concluir que asume un papel independiente y que constituye la esencia del cambio filosófico producido por Kolmogorov en la teoría de la probabilidad, consistente en la sustitución de la condición de equiprobabilidad y los resultados de sucesiones en condiciones repetibles por la adopción de una interpretación propensionalista del principio de Cournot como nexo principal con el mundo real.

Esta conclusión permite desmentir el argumento adoptado por la literatura de que el principal logro de los *Grundbegriffe* consistió en definir una axiomática basada en una idea primitiva de la probabilidad, independiente de cualquier interpretación acerca de la naturaleza del azar.

4.4. LA ADITIVIDAD NUMERABLE

En el primer capítulo de los *Grundbegriffe* Kolmogorov trata el problema de la aditividad finita y en el segundo capítulo a los primeros cinco axiomas adicionales el teorema de continuidad que, como se vio, introduce la condición de la aditividad numerable¹⁷. Ahora bien, dado que (a pesar de que el argumento original empleado por Fréchet para demostrar la σ -aditividad puede ser deducido de la definición empírica propuesta en Fréchet y Halbwachs, 1924, esta condición trasciende lo que puede ser verificado empíricamente, Kolmogorov manifiesta algunas reservas sobre su validez y la justifica exclusivamente en virtud de su utilidad en ciertos ámbitos de la investigación:

Como el nuevo axioma sólo es esencial en ámbitos infinitos de probabilidad, contrariamente a lo que ocurre con los cinco primeros axiomas, es casi imposible adivinar su significado empírico. En la descripción de cualquier proceso estocástico observable sólo podemos obtener ámbitos finitos de probabilidad. Los campos infinitos de probabilidad sólo se presentan en modelos ideales de procesos estocásticos reales. No obstante, tomaremos en consideración los modelos que satisfacen el Axioma 6, el cual resultó útil en ámbitos de investigación muy variados (Kolmogorov 1992 15).

¹⁷ Como se vio en la sección 2.1., la aditividad numerable fue introducida en la matemática pura por Borel (1909a, 1909b, 1912, 1925, 1930, 1939), quien la convirtió en el tema central de la teoría de la probabilidad al demostrar su ley fuerte de los grandes números.

4.5. LOS PROCESOS ESTOCÁSTICOS

La revisión de la perspectiva histórica de los *Grundbegriffe* permitió concluir que la filosofía de Kolmogorov fue menos influyente que sus axiomas, posiblemente debido a la limitación de la interpretación propensionalista respecto a la generalización de su *pensamiento probabilístico* a la teoría de los procesos estocásticos. Debe tenerse en cuenta que la primera y fundamental condición en el esquema de Kolmogorov es la repetibilidad ilimitada del sistema de condiciones \mathcal{G} y que, si bien este precepto se satisface cuando se define estrictamente un proceso estocástico en términos de sus probabilidades de transición, la cuestión se vuelve más comprometida cuando se analiza la distribución de probabilidades para los conjuntos de las trayectorias posibles, es decir, en los casos en los cuales se cuenta con una única trayectoria no repetible y, en consecuencia, solo es posible acceder a una definición débil del proceso. Kolmogorov logró resolver parcialmente esta insuficiencia demostrando que, en particular, un proceso de Markov discreto en el dominio del tiempo puede ser definido a partir de probabilidades simples, pero no logró extender este resultado a procesos continuos en el dominio del tiempo¹⁸.

5. CONCLUSIONES

Un análisis cuidadoso del contenido de los *Grundbegriffe* —obra que constituye la culminación natural de la interpretación del concepto de probabilidad a partir de los argumentos que proporcionan la teoría de la medida y la teoría de grupos— permitió arribar a las siguientes conclusiones sobre el pensamiento probabilístico de Kolmogorov: 1) Kolmogorov logró una formalización rigurosa del

¹⁸ Véase Landro y González (2009).

concepto de independencia estocástica como elemento diferenciador fundamental entre la probabilidad y la teoría de la medida; 2) según el principio de Cournot y de la axiomática de Von Mises, Kolmogorov asumió, en principio, una interpretación frecuentista de la probabilidad; 3) dada la imposibilidad epistemológica de poder asegurar el cumplimiento de las condiciones de equivalencia perfecta y de invariancia de los infinitos factores que afectan a las repeticiones que constituyen el fundamento de la axiomática frecuentista, Kolmogorov transitó hacia una interpretación posteriormente conocida como propensionalista, sobre la cual formuló su axiomática; 4) en consecuencia, la afirmación de la literatura de que la axiomática de Kolmogorov se basa en una idea primitiva de la probabilidad, independiente de cualquier interpretación acerca de la naturaleza del azar, es incorrecta; 5) a partir de los postulados del teorema de continuidad y la consiguiente inclusión del sexto axioma, proporcionó una justificación formal de la muy controvertida condición de aditividad numerable; 6) la imposibilidad de la generalización de sus interpretaciones de la noción de probabilidad a la definición débil de los procesos estocásticos continuos en el dominio del tiempo hizo que el contenido filosófico de los *Grundbegriffe* fuera menos influyente que su axiomática.

REFERENCIAS

- Bernoulli, Johann. *Ars conjectandi*. Thurnisiorum, 1713.
- Bernstein, Felix. “Über eine Anwendung der Mengenlehre auf ein aus der Theorie der Säkularen störungen herrührendes Problem”. *Mathematische Annalen* 71 (1912): 427-439. <<https://doi.org/10.1007/BF02216580>>
- Borel, Emile. *Éléments de la théorie des probabilités*. Hermann, 1909a.
- _____. “Les probabilités dénombrables et leurs applications arithmétiques”. *Rendiconti del Circolo Matematico di Palermo* 27 (1909b): 247-271. <<https://doi.org/10.1007/BF03019651>>

- _____. “Sur un problème des probabilités relatif aux fractions continues”. *Oeuvres de Emile Borel*, vol. 2. Centre National de la Recherche Scientifique, 1912. 1085-1091.
- _____. *Principes et formules classiques du calcul des probabilités*. Gauthier-Villars, 1925.
- _____. “Sur les probabilités universellement négligeables”. *Comptes Rendus Hebdomadaires de Séances de l'Académie des Sciences* 190 (1930): 537-559.
- _____. *Valeur pratique et philosophie des probabilités*. Gauthier-Villars, 1939.
- Buffon, Georges-Louis. “Essai d'arithmétique morale”. *Supplément à la Histoire Naturelle* 4 (1777): 46-148.
- Cantelli, Francesco P. “La tendenza ad un limite nel censo del calcolo delle probabilità”. *Rendiconti Circolo Matematico di Palermo* 41 (1916a): 191-201.
- _____. “Sulla legge dei grandi numeri”. *Atti Reale Accademia Nazionale dei Lincei, Memorie della Classe di Scienze Fisiche, Matematiche e Naturali* 11 (1916b): 330-349.
- _____. “Sulla probabilità come limite della frequenza”. *Atti Reale Accademia Nazionale dei Lincei, Memorie della Classe di Scienze Fisiche, Matematiche e Naturali* 26 (1917a): 39-45.
- _____. “Su due applicazioni d'un teorema di G. Boole alla statistica matematica”. *Atti Reale Accademia Nazionale dei Lincei, Memorie della Classe di Scienze Fisiche, Matematiche e Naturali* 26 (1917b): 39-45.
- _____. “Una teoria astratta del calcolo delle probabilità”. *Giornale dell'Istituto degli Attuari* 3 (1932): 257-265. <<https://doi.org/10.1122/1.2116457>>
- _____. “Consideration sur la convergence dans le calcul des probabilités”. *Annales de l'Institut Henri Poincaré* V (1935): 1-50.
- Cournot, Antoine Augustin. *Exposition de la théorie des chances et des probabilités*. Hachette, 1843.
- D'Alembert, Jean le Rond. “Doutes et questions sur le calcul des probabilités”. *Mélanges de Littérature d'Histoire et de Philosophie* 5 (1767): 277-278.
- _____. “Reflexions sur le calcul des probabilités”. *Opuscule mathématiques* 2 (1761): 1-25.

- Daniell, Percy J. "Integral products and probability". *American Journal of Mathematics* 43 (1921): 143-162. <<https://doi.org/10.1097/00000441-192107000-00044>>
- Daston, Lorraine J. "D'Alembert's critique of probability theory". *Historia Mathematica* 6 (1979): 259-279. <[https://doi.org/10.1016/0315-0860\(79\)90125-3](https://doi.org/10.1016/0315-0860(79)90125-3)>
- _____. "How probabilities came to the objective and subjective". *Historia Mathematica* 21 (1994): 235-254. <<https://doi.org/10.1179/tav.1994.1994.2.235>>
- De Scheemaerkere, Xavier. y Szafarz, Ariane. *The special status of mathematical probability: A storical sketch*. Centre Emile Bernheim, Research Institute in Management Sciences, 2008.
- Doob, Joseph Leo. "Kolmogorov's early work on convergence theory and foundations". *Annals of Probability* 17 (1989): 815-821.
- _____. "The development of rigor in mathematical probability, 1900-1950". *American Mathematical Monthly* 103 (1996): 586-595. <<https://doi.org/10.1080/00029890.1996.12004791>>
- Faber, Georg. "Über stetige Funktionen". *Mathematische Annalen* 69 (1910): 372-443. <<https://doi.org/10.1007/BF01456327>>
- Fréchet, Maurice. "Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements". *Rendiconti Reale Istituto Lombardo di Scienze et Lettere* 63 (1930a): 1059-1062.
- _____. "Sur la convergence 'en probabilité'". *Metron* 8 (1930b): 1-50.
- _____. *Recherches théoriques modernes sur la théorie des probabilités*. Gauthier-Villars, 1937-1938.
- Fréchet, Maurice y Halbwachs, Maurice. *Le calcul des probabilités à la portée de tous*. Dunod, 1924.
- Hadamard, Jacques. "Les principes du calcul des probabilités". *Revue de Métaphysique et de Morale* 39 (1922): 114-126.
- Hausdorff, Felix. *Grundzüge der Mengenlehre*. Von Weit, 1914.
- Hawkins, Thomas. *Lebesgue's theory of integration: Its origins and development*. University of Wisconsin Press, 1970.
- Hilbert, David. "Mathematical problems". *Bulletin of the American Mathematical Society* 8 (1902): 213-237.

- Holgate, Philip. "Independent functions: Probability and analysis in Poland between wars". *Biometrika* 84 (1997): 161-173.
- Khintchine, Aleksandr Y. "Über einen satz der Wahrscheinlichkeitsrechnung". *Fundamenta Mathematicæ* 6 (1924): 9-20.
<<https://doi.org/10.4064/fm-6-1-9-20>>
- _____. "Sur les lois de grandes nombres". *Comptes Rendus Academie des Sciences* 188 (1929): 477-480.
- _____. "Remarques sur les suites d'événements obeisant à la loi des grandes nombres". *Matematiskii Sbornik* 39 (1932a): 115-119.
- _____. "Sulle sucessioni stazionarie di evento". *Giornale dell'Istituto Italiano degli Attuari* 3 (1932b): 267-272.
- Kolmogorov, Andrei N. "The general theory of measure and the calculus of probability". *Collected works of the Mathematical Section, Communist Academy, Section for Natural and Exact Sciences* 1 (1929): 8-21.
- _____. "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung". *Mathematische Annalen* 104 (1931): 415-458.
<<https://doi.org/10.1007/BF01457949>>
- _____. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, 1933.
- _____. "Obituary for Evgen Evgenevich Slutsky". *Russian Mathematical Surveys* 3 (1948): 143-151.
- _____. "On tables of random numbers". *Sankhya* 25 (1963): 369-376.
- Laemmel, Ralf. *Untersuchungen über die Ermittlung von Wahrscheinlichkeiten*. University of Zurich, 1904.
- Landro, Alberto H. y González, Mirta L. *Elementos de econometría de los fenómenos dinámicos*. Ediciones Cooperativas, 2009. <http://bibliotecadigital.econ.uba.ar/download/libros/Landro-Gonzalez_Elementos-econometria-fenomenos-dinamicos-1.pdf>.
- _____. *Acerca del problema de Bernoulli y la determinación del verdadero valor de una probabilidad*. Ediciones Cooperativas, 2016. <http://bibliotecadigital.econ.uba.ar/download/libros/Landro-Gonzalez_Acerca-del-problema-de-Bernoulli-2016.pdf>.
- Lebesgue, Henri. "Sur une généralization de l'intégrale définie". *Comptes Rendus Académie des Sciences* 132 (1901): 1025-1028.

- _____. *Leçons sur l'intégration et la recherche des fonctions primitives*. Gauthier-Villars, 1904.
- Lévy, Paul. *Calcul des probabilités*. Gauthier-Villars, 1925.
- Łomnicki, Zbigniew. "Nouveaux fondements du calcul des probabilités (definition de la probabilité fondée sur la théorie des ensembles)". *Fundamenta Mathematica* 4 (1923): 34-71. <<https://doi.org/10.4064/fm-4-1-34-71>>
- Łomnicki, Zbigniew y Ulam, Stanislaw. "Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités. I: Variables indépendantes". *Fundamenta Mathematica* 23 (1934): 237-278. <<https://doi.org/10.4064/fm-23-1-237-278>>
- Nikodym, Otto. "Sur une generalization des integrals de M. J. Radon". *Fundamenta Mathematica* 15 (1930): 131-179. <<https://doi.org/10.4064/fm-15-1-131-179>>
- Pier, Jean Paul. "Intégration et mesure 1900-1950". *Development of mathematics 1900-1950*, Jean Paul Pier (ed.). Birkhäuser, 1994a. 517-564.
- _____. (ed.). *Development of mathematics 1900-1950*. Birkhäuser, 1994b.
- Popper, Karl R. "Probability magic or knowledge out of ignorance". *Dialectica* 11 (1957a): 354-374.
- _____. "The propensity interpretation of probability". *British Journal for the Philosophy of Science* 10 (1957b): 25-42.
- _____. *The logic of scientific discovery*. Hutchinson, 1959.
- Prokhorov, Yuri V. "Convergence of random processes and limit theorems of probability theory". *Theory of Probability and its Applications* 1 (1956): 157-214.
- Radon, Johann. "Theorie und Anwendungen der absolut asdditiven Mengenfunktionen". *Akademie der Wissenschaften in Wien, Sitzungsberichte, Mathematisch-naturwissenschaftliche Klasse, Abteilung* 122 (1913).
- Sierpiński, Waclaw. "Sur une définition axiomatique des ensembles mesurables". *Bulletin International Academy of Sciences Serie A* (1918): 173-178.
- Slutsky, Yevgueni. "On the question of the logical foundation of the theory of probability". *Bulletin of Statistics* 12 (1922): 118-119.
- _____. "Über stocastische Asymptoten und Grenzwerte". *Metron* 5 (1925): 3-89.

- Steinhaus, Hugo. “Les probabilités dénombrables et leur rapport à la theorie de la mesure”. *Fundamenta Mathematica* 4 (1923): 286-310.
<<https://doi.org/10.4064/fm-4-1-286-310>>
- _____. “Über die Wahrscheinlichkeit dafür daß der Konvergenzkreis einer Potenzreihe ihre natürliche Grenze ist”. *Athematische Zeitschrift* 31 (1930a): 408-410.
- _____. “Sur la probabilité de la convergence de séries”. *Studia Mathematica* 2 (1930b): 21-39.
- Ulam, Stanislaw. “Zum Massbegriffe in Produkträumen”. *Verhandlung des Internationalen Mathematiker-Kongress, Zurich* 2 (1932): 118-119.
- Von Kries, Johannes. *Die Principen der Wahrscheinlichkeitsrechnung. Eine logische Untersuchung*. Mohr, 1886.
- Von Mises, Richard. *Wahrscheinlichekeitsrechnung und ihre Anwendung in der statistic und teoritischen Physic*. Rosenberg, 1931.
- Von Plato, Jan. *Creating modern probability*. Cambridge University Press, 1994. <<https://doi.org/10.1017/CBO9780511609107>>
- Wiman, Anders. “Über eine Wahrcheinlichketaufgabe bei Kettebrucentwicklungen”. *Öfverseghr af Kongliga Svenska Vatenkaps-Akademiens Förhandlingar Fentindesjunde Ärgangen* 57 (1900): 829-841.
- _____. *Bemerkung über eine von Glydén Aufgeworfene Wahrscheinlichkeitsfrage*. HÅkan Ohlssons Boktykeri, 1901.

ELEMENTS FOR AN EPISTEMOLOGICAL REVISION OF KOLMOGOROV'S AXIOMATICS*

ELEMENTOS PARA UNA REVISIÓN EPISTEMOLÓGICA DE LA AXIOMÁTICA DE KOLMOGOROV

ALBERTO LANDRO**

Centro de Investigaciones en Econometría -
Universidad de Buenos Aires y Departamento de
Investigaciones - Universidad del CEMA
Ciudad de Buenos Aires, Argentina.
alandroar@yahoo.com.ar
<https://orcid.org/0009-0000-4402-5726>

MIRTA L. GONZALEZ

Centro de Investigaciones en Econometría -
Universidad de Buenos Aires; Universidad del
CEMA y Gerencia de Estudios Económicos del
Banco Central de la República Argentina
Ciudad de Buenos Aires, Argentina.
mirtagonzalezar@yahoo.com.ar
<https://orcid.org/0009-0001-1186-6104>



ABSTRACT

Although the literature on probability has traditionally considered certain assumptions about what were the aims and scope of the "Grundbegriffe der Wahrscheinlichkeitsrechnung" in the development of probability theory, a detailed analysis of Kolmogorov's work in particular regarding its passage through the frequentist and propensionalist interpretations, leads to the conclusion that hypotheses usually raised

* This article should be quoted: Landro, Alberto & Gonzalez Mirta L. "Elements for an Epistemological Review of Kolmogorov Axiomatics". *Revista Colombiana de Filosofía de la Ciencia* 24.48 (2024): 109-136. <https://doi.org/10.18270/rfc.4412>

** The opinions expressed in this article are those of the authors and do not reflect the views of the institutions.

in the literature about its axiomatics being based on a primitive concept of probability, independent of any interpretation about the nature of chance, are generally inaccurate and that these errors in interpretation led to inadequate conceptual and epistemological approaches.

Keywords: epistemology of probability; probability theory; propensionalist-frequentist interpretation of the concept of probability; Kolmogorov's axiomatics; measure theory; group theory.

RESUMEN

La literatura sobre la probabilidad ha considerado tradicionalmente ciertos supuestos acerca de cuáles fueron los objetivos y alcances de los Grundbegriffe der Wahrscheinlichkeitsrechnung en el desarrollo de la teoría de la probabilidad. Sin embargo, un análisis detallado de la obra de Kolmogorov, en particular en lo referido a su transcurso por las interpretaciones frecuentista y propensionalista, permitió concluir que las hipótesis planteadas habitualmente en la literatura acerca de que su axiomática se basa en un concepto primitivo de la probabilidad, independiente de cualquier interpretación acerca de la naturaleza del azar, son en general inexactas y que estos desaciertos en la interpretación condujeron a planteos conceptuales y epistemológicos inadecuados.

Palabras clave: epistemología de la probabilidad; teoría de la probabilidad; interpretación propensionalista-frecuentista del concepto de probabilidad; axiomática de Kolmogorov; teoría de la medida; teoría de grupos.

1. DIE GRUNDBEGRIFFE DER WAHRSCHEINLICHKEITSRECHNUNG

According to the literature on probability, the aims of Andrei Nikolaievitch Kolmogorov's proposal, contained in the *Grundbegriffe* (1933), were (1) to solve the problem of the lack of rigorousness with which mathematicians of the 18th, 19th, and first half of the 20th centuries treated the problems of probability; 2) to consider probability as a primitive idea, independent of any interpretation of the nature of chance; and 3) starting from the conception of probability theory as a particular case of an abstract mathematical theory, to define an axiomatic that would act as a reference system for the axiomatics inherent in the different interpretations of the nature of chance.

The careful analysis of the evolution of Kolmogorov's probabilistic thinking carried out in this paper allowed us to conclude that some of these explicit or implicit assessments of the literature commonly used in undergraduate courses on the *Grundbegriffe*'s theoretical-conceptual scopes are, in general, debatable and that their interpretations of the philosophical foundations of axiomatics affected the scopes of the derived theorems.

2. AN ANALYSIS OF THE BACKGROUND

2.1. THE MEASURE THEORY

The first attempts to link set measure theory and Lebesgue integration with probability quantification were proposed by Borel, initiating the process that culminated in the axiomatic system contained in the *Grundbegriffe*.

By formulating what is known as the strong law of large numbers, Borel (1909a, 1909b) succeeded in formalizing the relationship between the theory of measurement and a geometrical interpretation that forms the core of classical probability theory. But the theoretical justification of this relationship, using generalizations to numerable sets of the total probability and compound probability theorems, is undeniably incomplete due to circumstances possibly attributable to philosophical reasons related to his refusal to accept the notion of numerable probability, which forced him to employ frequentist arguments that consider the probabilities to be limits of successions of independent repetitions.

It was Faber (1910) and Hausdorff (1914) who amended this limitation, from which they obtained rigorous proofs of the strong law of large numbers using exclusively measurement theory arguments, generalized, as will be seen later, by Cantelli (1916a, 1916b, 1917a)¹. The result of these proposals was a strong law of large numbers interpretable according to two versions: one related to the domain of real numbers and the other to probability.

Following the canons of Sierpiński's axiomatization (1918) and Lebesgue's measure (1901, 1904)², Steinhaus (1923, 1930a, 1930b) in his work on theorems in the limit proposed a solution to this conceptual duality by means of the axiomatization of Borel's theory³:

¹ Cantelli's results were of great importance in that they prompted other authors to develop different definitions of stochastic convergence (see Doob 1989, 1996; Von Plato 1994).

² E In his axiomatic expression of measure theory, Sierpiński (1918) characterized the Lebesgue class of measurable sets as the smallest class C^* of sets satisfying the following conditions: axiom 1: every set $E \subset C$ can be assigned a real number $p(E)$ that defines its measure and satisfies the axioms that follow; axiom 2: every finite and closed interval $E \subset C^*$ possesses a length and a measure; axiom 3: the class C^* is closed for any finite and numerable union of unions of disjoint elements and p is a finite and additive numerable measure; axiom 4: let E_1 and E_2 be two sets $E_1 \supset E_2$ and let $E_1, E_2 \subset C^*$, then, $E_1/E_2 \subset C^*$; axiom 5: if the set $E \subset C^*$ y $p(E) = 0$, then, every subset of $E^* \subset C^*$.

³ Steinhaus' 1923 memoir gave rise to the so-called Polish school in the study of independent functions, to which we owe important contributions to probability theory.

Let \mathcal{A} be the set of all infinite binary sequences of events (i.e. admitting only two possible outcomes), \mathcal{C} a class of subsets of \mathcal{A} and a real function p representing the probabilities of the elements of a subset E of \mathcal{C} , *axiom 1*: $p(E) \geq 0$; *axiom 2*: a) for any finite binary sequence E^* , the subset E of \mathcal{A} consisting of all infinite binary sequences starting with E^* belongs to \mathcal{C} ; b) if two such finite binary sequences, E_1^* and E_2^* , differ by only one element then $p(E_1) = p(E_2)$ (where E_1 y E_2 denote the subsets of \mathcal{A} starting with the sequences E_1^* y E_2^* , respectively); c) $p(\mathcal{A}) = 1$; *axiom 3*: the class \mathcal{C} is closed with respect to finite and countable unions of disjoint elements, and the function p satisfies the conditions of finite additivity and countable additivity; *axiom 4*: if $E_1 \supset E_2$ and $E_1, E_2 \in \mathcal{C}$, then, $(E_2 / E_1) \in \mathcal{C}$; *axiom 5*: if $E \in \mathcal{C}$ and $p(E) = 0$, then for any subset E^* such that $E \supset E^*$, $p(E^*) = 0$.

Steinhaus' axiomatics includes Sierpiński's axioms 1, 3, 4 and 5 and an axiom (which Steinhaus proved to be equivalent to his axiom 2), according to which the probability that event E is in an interval J ($p(E \subset J)$) is equal to the size of the interval⁴.

Subsequently Khintchine (1929, 1932a, 1932b), Kolmogorov (1929, 1931, 1933) and Prokhorov (1956), among others, began the succession of generalizations of the necessary and sufficient conditions to ensure the validity of the law of large numbers⁵.

Kolmogorov (1933) established that the difference between probability and measurement theory is the condition of independence. This fundamental postulate can be considered the starting point '[...] towards a meaningful consensus about the mathematical aspects of probability [...] as

⁴ See Holgate (1997).

⁵ See Landro y González (2016).

its philosophical interpretation runs towards the alternative of probability being epistemic or objective' (De Schewemaerkere & Szafarz 2008 2)⁶.

2.2. GROUP THEORY

Another important antecedent in the process of constructing the Grundbegriffe and thus Kolmogorov's formulation of axiomatics was the work of Slutsky (1922): "Slutsky was the first to give a correct version of the purely mathematical content of probability theory" (Kolmogorov 1948 143). His proposal took as its starting point the substitution of the term *probability* for that of *valence*⁷. In his concept, probability would only be a representation of the valence calculus. He replaced the concept of equiprobability with the simple assumption that if certain cases to which a valence α is assigned are subdivided, the valences assigned to the subcases should sum to α (the valences assigned to the subcases may or may not be equal). In this way, the theorems of total probability and compound probability became part of a calculus (which conceptually cannot be considered a probability calculus) as abstract as group theory.

In his 1929 memoir, Kolmogorov adopted Slutsky's (1925) approach of conceiving of probability as a particular case of an abstract theory and proposed an interpretation of probability related to a function $M(\cdot)$ (which he called the *measurement specification*) which assigns non-negative numbers $M(E)$ to each element E of a class of subsets of a set of attributes \mathcal{A} ; furthermore, he assumed only that given two disjoint elements E_1 y E_2 disjoint, then

⁶ The development of Lebesgue's theories of measure and integration in the late nineteenth and early twentieth centuries has already been extensively studied (see, for example, Hawkins, 1970; Pier, 1994a, 1994b). Therefore, only a brief review of those works that directly influenced the content of the *Grundbegriffe* will be attempted here.

⁷ Laemmel (1904) had already used the German term *valenz*.

the function $M(\cdot)$ shall assign a number to any two of the sets E_1 y E_2 , so that the number to be assigned to the third set $E_1 \cup E_2$ is determined by addition, it being verified that⁸

$$M(E_1 \cup E_2) = M(E_1) + M(E_2).$$

With respect to the probability of occurrence of E_2 conditional on the assumed occurrence of E_1 , Kolmogorov proposed the following definition, based on the results obtained by Steinhaus:

$$M(E_2 / E_1) = \frac{M(E_1 \cap E_2)}{M(E_1)}.$$

In 1930, as a logical extrapolation to the case of infinite successions of events, Kolmogorov proved that the only probabilities satisfying Steinhaus' axioms are those obtained from a system isomorphic with the Lebesgue measure in the interval $[0, 1]$ and, consequently, with Sierpiński's axiomatics. His extension of the scope of these axioms included all successions of random variables, even complex variables, but did not transcend the geometrical context and thus failed to formulate a notion of probability in abstract spaces. The task was given to the mathematicians Stanislaw Ulam and Zbigniew Łomnicki, who proved in their memoirs of 1923, 1932 and 1934 that $M(\cdot)$ is a probability measure on a complete σ -algebra, and that, from a numerable succession of spaces, with probability measures M , one can construct a probability measure satisfying the same conditions on the space defined by the Cartesian product of the marginal event spaces.

⁸ Note that these conditions coincide with axioms 3 and 4 of Steinhaus (1923).

The fact that operations between events can be identified with operations between sets in such a way that the formal properties of the latter remain, allows us to remove any material basis from events and to consider them as mathematical entities susceptible of applying to them procedures, reasoning and deductions proper to mathematical logic. However, beyond their undeniable importance, these results on the progress of the justification of probability theory allow us to conclude that, although the representation of events by means of sets constitutes a useful device to facilitate their mathematical treatment, it does not allow any interpretation of their nature. In other words, an abstract mathematical structure cannot grasp the intuitive content of the concept of event.

2.3. FREQUENCY ABSTRACTION

For his part, Cantelli (1917b, 1932) introduced an abstract theory of probability in which he discards the use of notions linked to empirical questions such as those of *possibility, event, probability and independence*, and in which he only formally retains the classical rules of total and compound probability. Denoted by $m(E)$ the measure of a set E , he established that: i) given two disjoint subsets E_1 y E_2 it is verified that:

$$M(E_1 \cup E_2) = M(E_1) + M(E_2);$$

ii)

$$0 \leq \frac{m(E_1 \cap E_2)}{m(E_i)} \leq 1 \quad (i = 1, 2),$$

in particular if it is verified that $m(E_1 \cap E_2) = m(E_1)m(E_2)$, consider that the sets E_1 y E_2 are '*multiplicative*'; and iii) even Bernoulli's law of large numbers and its generalizations and Khinchin's law⁹ of the iterated logarithm can be treated at the same level of abstraction as his theory proposes.

In response to Slutsky's consideration that since the empirical sequences of independent repetitions possess properties that transcend the existence of a limit, probability could not be assimilated to the limit of a relative frequency, Cantelli (1916a, 1916b, 1917a, 1935) independently developed a project of justification of the frequentist interpretation of probability according to his abstract axiomatic system. According to his approach, the classical probability calculus - which considers the particular class of events that are characterised by the fact that the physical mechanism that generates randomness symmetrically includes all possible outcomes - should be developed (in a circular fashion) in three stages: the first was to experimentally test the assumption of equiprobability (by checking whether they occur with the same frequency) and thus empirically justify the use of the rules of total and compound probability; the second stage was aimed at developing, without reference to its empirical justification, an abstract theory based solely on its axiomatics; and the third was to derive probabilities from such a theory and use them to predict relative frequencies.

⁹ As a generalization of the law of large numbers Kintchine (1924) proved the law of the iterated logarithm, according to which, given a succession of random variables $Y_i: b(I, \theta)$, exists a function: $f(n) = \sqrt{2n\theta(1-\theta)\log[\log(n)]}$ characterizing the rate of convergence, such that, for all ε, δ , a value n_0 and a variable $Y_{(n)} = \sum_{i=1}^n Y_i$, it is verified that $p\left(\left|\frac{Y_{(n)}}{f(n)}\right| < 1 + \varepsilon\right) = 1 - \delta$ (see Landro; González (2016)).

3. AXIOMATICS

As a culmination of the development of measurement theory between 1900 and 1930, *Grundbegriffe der Wahrscheinlichkeitsrechnung* was published:

It was precisely when Mr. Borel introduced this new kind of [numerable] additivity into the probability calculus - in 1909 - that all the elements necessary to explicitly formulate the complete body of axioms of (modernized classical) probability theory were summed up. It is not enough to have all the ideas 'in mind', one must be sure that this totality of ideas is sufficient, express them together and take the responsibility of affirming that nothing else is necessary to build the theory. This is what Mr. Kolmogorov did. This was his achievement (Fréchet 1930a 1059).

In his 62 pages, Kolmogorov managed to introduce all the mathematics synthesized mainly in the work of Borel, Steinhaus, Slutsky, Cantelli, and Daniell (in particular as regards their consistency theorem, i.e., the construction of probabilities in infinite-dimensional spaces), Radon (1913) and Nikodym (1930) (as regards conditional probabilities and expected values), and, in particular, the theory of Fréchet (1930b) (as regards the stochastic independence condition)¹⁰.

In response to the sixth of Hilbert's famous collection of problems, according to which 'We must treat by means of axioms those physical sciences in which mathematics plays an important role, in particular, the theory of

¹⁰ It is worth mentioning that the core of the *Grundbegriffe* bibliography includes authors such as Sergei Bernstein, Emile Borel, Paul Lévy, Antoni Lomnicki, Evgeny Slutsky, Hugo Steinhaus, Richard von Mises, Guido Castelnuovo and Maurice Fréchet. To this list should be added the contributions of Henri Lebesgue, Waclaw Sierpiński, Constantin Carathéodory, Johann Radon, Otton Nikodym, Percy Daniell and Norbert Wiener, whose work constitutes the philosophical and mathematical heritage that Kolmogorov received.

probability and mechanics' (1902 237), Kolmogorov adopted a philosophical position very similar to the justification used by Fréchet regarding the need to make the concept of probability independent of the geometrical interpretation of Borel, Lévy, and Slutsky. This transition from the geometric interpretation to that proposed by Fréchet was considered by Kolmogorov to be a fundamental basis of probability theory and formed the argument for the formulation of what many probabilists arbitrarily call the classical axiomatics of probability¹¹:

- *Axiom 1*: the events form a σ -algebra s , i.e. a closed class with respect to the operations of union, intersection and negation of numerable sets of events and of the limit of sequences of events that is: (a) if events $E_j \in s$ ($j = 1, 2, \dots$), then, $\bigcup_{j=1}^{\infty} E_j \in s$; and b) if $E_j \in s$ ($j = 1, 2, \dots$), then, $\bigcap_{j=1}^{\infty} E_j \in s$ (in fact, given that $E_i \cap E_j = (E_i \cup E_j) - [(E_i - E_j) \cup (E_j - E_i)]$, this property can be considered as a consequence of postulate a); and c) given a succession of events, this property can be considered as a consequence of postulate a) $\{E_1, E_2, \dots\}$, belonging to s , then $\lim_{j \rightarrow \infty} E_j \in s$.
- *Axiom 2*: given a sample space Ω , it is verified that $\Omega \in s$.
- *Axiom 3*: associated with each event $E \in s$, there exists a non-negative real number, $p(E)$, which will be referred to as *probability of occurrence of the event E*.

¹¹ This consideration is evident in Kolmogorov's apology in the preface to the *Grundbegriffe* for the results obtained by Fréchet: 'After Lebesgue's investigations, the analogies between the measure of a set and the probability of an event and the integral of a function and the mathematical expectation of a random variable became clear. These analogies could be extended since, for example, many properties of independent random variables are analogous to the corresponding properties of orthogonal functions. However, to base probability theory on these analogies, it was necessary to make the theory of measurement and integration independent of the geometrical elements that still figured prominently in Lebesgue's proposal. This achievement was obtained by Fréchet'. (1992 p.v).

- *Axiom 4*: the probability of occurrence of at least one of the events included in the sample space is equal to 1: $p(\Omega) = 1$.
- *Axiom 5 (additivity)*: let them be E_1 y E_2 two incompatible events, i.e., such that they cannot occur simultaneously ($E_1 \cap E_2 = \emptyset$), then it shall be verified that

$$p(E_1 \cup E_2) = p(E_1) + p(E_2)$$

(the first strict formulation of this principle is found in J. Bernoulli's *Ars conjectandi* (1713)).

- *Axiom 6 (continuity)*: given a monotonic succession of events $E_i \supset E_{i+1}$ ($i = 1,2,3, \dots$), can be written:

$$E_i = E_1 \cup (E_1 \cap \bar{E}_2) \cup \dots \cup (E_{i-1} \cap \bar{E}_i) \quad (i = 2,3, \dots),$$

it follows that

$$p(E_i) = p(E_1) + \sum_{j=2}^i p(E_{j-1} \cap \bar{E}_j) \quad (j = 2,3, \dots),$$

so that, applying the limit operator member by member, will be

$$\begin{aligned} \lim_{i \rightarrow \infty} p(E_i) &= p(E_1) + \sum_{j=2}^{\infty} p(E_{j-1} \cap \bar{E}_j) = p \left[E_1 \bigcup_{j=2}^{\infty} (E_{j-1} \cap \bar{E}_j) \right] = \\ &= p \left\{ \lim_{i \rightarrow \infty} [E_1 \cup_{j=2}^i (E_{j-1} \cap \bar{E}_j)] \right\} = p \left(\lim_{i \rightarrow \infty} E_i \right). \end{aligned}$$

Thus, it can be concluded that the probability is a continuous function with respect to any monotonic succession of events.

Likewise, if the sequence is such that $E_i \supseteq E_{i+1}$ ($i = 1,2, \dots$), then it shall be verified that $\bar{E}_i \subseteq \bar{E}_{i+1}$ and, therefore, given the existence of the limit of the succession of the $p(\bar{E}_i)$, it will be

$$\lim_{i \rightarrow \infty} p(\bar{E}_i) = 1 - \lim_{i \rightarrow \infty} p(\bar{E}_i) = 1 - p(\cup_{i=1}^{\infty} \bar{E}_i) = p(\cap_{i=1}^{\infty} E_i) = p\left(\lim_{i \rightarrow \infty} E_i\right).$$

then, given the first five axioms, it is easily shown that this sixth axiom of continuity is equivalent to the condition of *complete additivity* or *numerable additivity* or σ -*additivity* (which, obviously, contains simple additivity as a particular case): let be $\{E_1, E_2, \dots, E_n, \dots\}$ a set of pairwise incompatible events, i.e. such that $E_i \cap E_j = \emptyset$ ($i \neq j; i, j = 1, 2, \dots$). By induction, it is shown that

$$\cup_{j=1}^{n+1} E_j = \cup_{j=1}^n E_j + E_{n+1}.$$

Given that any two of the events E_j ($j = 1, 2, \dots, n$) and E_{n+1} are incompatible, it shall be

$$p(\cup_{j=1}^{n+1} E_j) = p(\cup_{j=1}^n E_j) + p(E_{n+1}) = \sum_{j=1}^{n+1} p(E_j).$$

In the same way, it can be written:

$$\cup_{j=1}^{\infty} E_j = (\cup_{j=1}^n E_j) \cup (\cup_{j=n+1}^{\infty} E_j),$$

and, as each of the events E_j ($j = 1, 2, \dots, n$) is incompatible with each of the events E_j ($j = n + 1, n + 2, \dots$), it shall be verified that

$$\begin{aligned} p\left(\bigcup_{j=1}^{\infty} E_j\right) &= p\left(\bigcup_{j=1}^n E_j\right) p\left(\bigcup_{j=n+1}^{\infty} E_j\right) = \\ &= \sum_{j=1}^n p(E_j) + p\left(\bigcup_{j=n+1}^{\infty} E_j\right) = \end{aligned}$$

$$= \lim_{n \rightarrow \infty} \sum_{j=1}^n p(E_j) + \lim_{n \rightarrow \infty} p(\bigcup_{j=n+1}^{\infty} E_j) = \sum_{j=1}^{\infty} p(E_j).$$

It should be noted that the events $(\bigcup_{n+1}^{\infty} E_j)$ form a decreasing sequence, i.e. that $(\bigcup_{n+1}^{\infty} E_j) \subset (\bigcup_n^{\infty} E_j)$, so that $\lim_{n \rightarrow \infty} (\bigcup_{n+1}^{\infty} E_j) = \emptyset$.

An analytical reading allows us to conclude, then, that Kolmogorov's axiomatic can be summed up in the principle that $p(\cdot)$ is a non-negative, additive function in the sense of Fréchet (1937-1938), with the addition of the condition $p(\Omega) = 1$.

As a corollary to these considerations, it can be concluded that the *Grundbegriffe* strictly fulfilled the aim stated in its preface: '[...] to fill a gap in the literature on probability theory (including in that literature Fréchet's book, which is in preparation) and to provide a presentation of the system in its entirety, free of superfluous complications' (Kolmogorov 1992 p.v).

4. A REVISION OF THE PHILOSOPHICAL CONTENT OF THE *GRUNDBEGRIFFE*

4.1. INDEPENDENCE CONDITION

The *Grundbegriffe* not only succeeded in synthesizing the development of measurement theory, but also made fundamental advances over the mathematics of Borel, Fréchet, Daniell and Cantelli: 1) in contrast to Cantelli's abstract system, Kolmogorov proposed an axiomatic associated with the concept of probability, which made it possible to develop, from the formulation of a general method of representation of the theory of measurement, an independ-

ent theory of probability conceived as a conceptually separate branch of mathematics; 2) in this sense, it incorporates the concept of stochastic independence of events as the fundamental differentiating factor between probability and measurement theory; and 3) at least in formal terms, the result obtained by Kolmogorov, insofar as it considers an arbitrary cardinality of the elements of the set, is more general than that provided by Daniell (1921), which considers a numerable cardinality - a generalization whose exclusively conceptual character allows us to conclude that it does not imply significant mathematical modifications, nor is it relevant in applications.

4.2. FREQUENCY TRANSITION

While Kolmogorov's axiomatics possesses the ability to transform the probability values of simple events into the probability values of complex events, insofar as it only determines the probability values 0 and 1 assigned to what he defines as impossible and certain events, respectively, it evidences its inability to assign probability values to non-trivial events such as *impossible* or *certain events*. The attempt to solve this limitation led Kolmogorov to move towards a frequentist interpretation and, subsequently, to the adoption of a definition in line with Popper's propensionalist interpretation.

The unequivocal circumstance that confirms this transition to the frequentist interpretation by Kolmogorov is his acceptance of Cournot's principle and Von Mises' axiomatics (1931), both of which postulate the approximation of the results of a set of repeatable conditions to the assumed true value of a probability as the main link of the axiomatics to the real world. This approach is summarized in the following paragraph of Section 2 of the *Grundbegriffe* (two pages long, under the heading '*Das Verhältniss zur Erfahrungswelt*')

It is possible to assume that an outcome A of eventual occurrence under certain conditions ζ , which we do not intend to discuss here, can be assigned a real number $p(A)$ that possesses the following properties: (a) it can be taken as practically certain that, if the system of conditions ζ is repeated a sufficiently large number of times (n) and it is verified that event A occurs m times, then the quotient $\frac{m}{n}$ will differ slightly from $p(A)$; b) if $p(A)$ is very small, then it can be taken as practically certain that event A will not occur in an individual repetition of conditions ζ (1992 4).

The so-called *Cournot principle* (1843) is related to the existence of events included in a domain Ω to which a unit probability is also assigned: these are the so-called *quasi-certain* or *certain-in-probability events*. The negation of a quasi-certain event defines an event that, not coinciding with the impossible event, behaves as such in terms of its probability and is called *quasi-impossible* or *impossible-in-probability*¹². These events possess the following properties: 1) the union of two impossible-in-probability events is an impossible-in-probability event (a property that can be generalised to any finite or infinite numberable set of events); 2) let E_I be an impossible event in-probability and E a neither certain nor impossible event, then it is verified that $(E_I \cap E)$ is an impossible event in probability and that $(E_I \cup E)$ is a neither certain nor impossible event; 3) let E_C be a certain event in-probability and E neither certain nor impossible event, then it is verified that $(E_C \cap E)$ is a neither certain nor impossible event and that $(E_C \cup E)$ is a certain event in-probability.

These definitions derive from the concepts of *morally impossible* and *morally certain* events originally formulated by Bernoulli (1713). *Probabilistic moral certainty* was widely discussed in the 18th century. D'Alembert (1761,

¹² That is, an event whose probability, while not effectively equal to zero, is of purely metaphysical interest.

1767, 1780) raised the need to distinguish between *metaphysically impossible events* (immune to the test of experience and *physically impossible*¹³). In this respect, Buffon (1777) considered the difference between moral and physical certainty to be a matter of degree of approximation; in his view, an event with probability $\frac{9999}{10000}$ is *morally true*, while an event with a much higher probability can be considered *physically true*.

Borel (1909a, 1912, 1925, 1930, 1939), often using a literary rather than a mathematical or philosophical style, adopted the treatment of events with infinitesimal probability as impossible and established a classification of them more refined than Buffon's, according to which a probability of 10^{-6} is negligible on a *human scale*, a probability of 10^{-15} is negligible on an *terrestrial scale*, and a probability of 10^{-50} is negligible on a *cosmic scale*¹⁴.

Cournot, building on Borel's concept of geometric probability, came up with a formulation of the law of large numbers according to which one can be sure that a morally impossible event will not occur (that its occurrence is mathematically possible but physically impossible) and, as a corollary, that it is mathematically impossible that, given a sufficiently long succession of repetitions, the relative frequency of a morally impossible event differs substantially from the null probability: the physically impossible event is, therefore, the one to which an infinitely small probability corresponds and this observation alone is sufficient to provide objective phenomenal substance to the mathematical theory of probability' (158)¹⁵.

¹³ See Daston (1979).

¹⁴ From Wiman's (1900, 1901) work on continued fractions, German mathematicians directed their efforts to show that there are sets of null measure and that these sets are *impossible* in probabilistic terms, which highlighted the gulf between null probability and infinitesimal probability (see Bernstein, 1912).

¹⁵ The expression phenomenal objective refers to the distinction made by Kant between *noumenon*, or thing-in-itself, and *phenomenon*, or object of experience. See Daston (1994).

Cournot's law of large numbers can be expressed in two ways: 1) a *strong* or *strict* form which considers that an event with an infinitesimally small probability will not occur in the next repetition and which, combined with the postulates of Bernoulli's theorem, allows us to conclude that the probability of an event can *always* be approximated by its relative frequency from a sufficiently long series of independent repetitions; and 2) a *weak* form which considers that an event with probability very close to zero will occur very rarely in a series of repetitions and which, combined with Bernoulli's theorem, allows one to conclude that the probability of an event can *usually* be approximated by its relative frequency from a sufficiently long series of independent repetitions.

Kolmogorov (according to the postulates of the *strong form* of the law of large numbers), following Cournot, assumed that the principle that assimilated an infinitesimally small probability to the impossibility of occurrence of an event provided an empirical meaning to classical probability, but did not include any consideration of the existence of the conceptual difference between the statement that an event of infinitesimal probability will not occur and the assumption that, by itself, that statement of non-occurrence of the event provides an empirical meaning to classical theory.

4.3. PROPENSIONALIST INTERPRETATION

The French and Russian probabilists who succeeded Laplace, in disagreement with what might be called the orthodox philosophy of his classical interpretation, began by assuming a subjectivist interpretation and later, from a misreading of the inverse version of *Bernoulli's theorem*¹⁶, of Cournot's principle

¹⁶ See Landro, González (2016).

and, without a valid conceptual basis of Bayes' theorem, attempted to establish the existence of objective probabilities of empirical significance by applying the law of large numbers.

To resolve these errors of interpretation, von Kries (1886) proposed, as a necessary condition for reasoning about objective probabilities of empirical significance from relative frequencies, the principle of equiprobability, which he called the *Spielraum principle*, according to which objective probabilities can be calculated from cases in which the circumstances that generate each case admit the same number of possible arrangements.

Regarding the assumption of equiprobability as a necessary condition, Hadamard (1922), following a purely deterministic position, argued that the notions of equally probable events (perfectly equivalent, according to his nomenclature) and of highly improbable events constitute the foundation of probability theory and that the fact that perfect equivalence is an unverifiable mathematical condition should not preclude empirical approximation.

In this respect, although Lévy (1925) agreed with Hadamard that the equiprobability condition can be considered a foundation of the mathematics of probabilities, he admitted that, given the empirical impossibility of its verification, if it is assumed as the sole basis of reasoning, the probabilities obtained will necessarily be subjective.

The Grundbegriffe review allowed us to conclude that, although it is undeniable that Fréchet's and Slutsky's proposals had a significant influence on Kolmogorov's thought, due to the limitations of Cournot's principle posed by Lévy, Kolmogorov formulated his axioms later, justified by Popper's propensionalist definition (1957a, 1957b, 1959). This consists of associating probabilities with the outcomes of a system of conditions \mathcal{C} repeatable (not specified in the text) and not with the collectives of the frequentist interpretation. Because of the adoption of this Popperian interpretation, the linking of Kolmogorov's abstract axiomatics to the world of experience materializes

based on non-operationalist principles, according to which fundamental theoretical concepts are usually assumed to be undefined and linked to observation only indirectly. In 'On tables of random numbers' (1963 369), Kolmogorov states that the limiting frequency is a theoretical idealization, a concept of purely mathematical interest:

The notion of probability as a limit of relative frequency does not contribute anything to the applicability of the results of probability theory to real practical problems that can only be dealt with from a finite number of repetitions.

On the other hand, based on Borel and Lévy's postulates, it was concluded here that the Kolmogorovian principle, according to which

[...] an outcome A of eventual occurrence under certain conditions ζ [...] can be assigned a number $p(A)$ such that a) it can be taken as practically certain that, if the system of conditions ζ is repeated a sufficiently large number of times (n) and it is verified that event A occurs m times, then the quotient $\frac{m}{n}$ will differ slightly from $p(A)$ (Kolmogorov 1992 4).

It can be deduced from the joint consideration of the principle: "b) if $p(A)$ is very small, then it can be taken as practically certain that event A will not occur in an individual repetition of the conditions ζ " (Kolmogorov 1992 4) and Bernoulli's theorem (which arises as a corollary of the axioms).

The conclusion that principle (a) mentioned in the previous paragraph not only precedes the axioms but is used for their formulation, and thus precedes Bernoulli's theorem, allowed, in turn, to conclude that it assumes an independent role and that it constitutes the essence of Kolmogorov's philosophical change in probability theory, consisting in the replacement of the

equiprobability condition and the results of sequences under repeatable conditions by the adoption of a propensionalist interpretation of Cournot's principle as the main link with the real world.

This conclusion disproves the argument adopted by the literature that the main achievement of the *Grundbegriffe* was to define an axiomatic based on a primitive idea of probability, independent of any interpretation of the nature of chance.

4.4. NUMERABLE ADDITIVITY

In the first chapter of the *Grundbegriffe* Kolmogorov deals with the problem of finite additivity and in the second chapter he adds to the first five axioms the continuity theorem which, as we have seen, introduces the condition of numerable additivity¹⁷. However, since (although the original argument used by Fréchet to demonstrate the σ - additivity can be deduced from the empirical definition proposed in Fréchet and Halbwachs, 1924, this condition transcends what can be verified empirically, Kolmogorov expresses some reservations about its validity and justifies it exclusively by virtue of its usefulness in certain fields of research:

Since the new axiom is only essential in infinite probability domains, contrary to the first five axioms, it is almost impossible to guess its empirical meaning. In the description of any observable stochastic process we can only obtain finite probability fields. Infinite probability fields only occur in ideal models of real stochastic processes. However, we will

¹⁷ As seen in section 2.1., numerable additivity was introduced into pure mathematics by Borel (1909a, 1909b, 1912, 1925, 1930, 1939), who made it the central theme of probability theory by proving his strong law of large numbers.

consider models that satisfy Axiom 6, which proved useful in a wide variety of research domains (Kolmogorov 1992 15).

4.5. STOCHASTIC PROCESSES

The review of the historical perspective of the *Grundbegriffe* led to the conclusion that Kolmogorov's philosophy was less influential than his axioms, possibly due to the limitation of the propensionalist interpretation regarding the generalization of his probabilistic thinking to the theory of stochastic processes. It should be noted that the first and fundamental condition in the Kolmogorov scheme is the unbounded repeatability of the system of conditions \mathcal{C} and that, while this precept is satisfied when a stochastic process is strictly defined in terms of its transition probabilities, the issue becomes more compromised when analyzing the probability distribution for the sets of possible trajectories, i.e., in cases where there is a single non-repeatable trajectory and, consequently, only a weak definition of the process is possible. Kolmogorov managed to partially solve this insufficiency by showing that, in particular, a discrete Markov process in the time domain can be defined from simple probabilities, but he did not manage to extend this result to continuous processes in the time domain¹⁸.

5. CONCLUSIONS

A careful analysis of the contents of the *Grundbegriffe*—the natural culmination of the interpretation of the concept of probability based on the arguments provided by measure theory and group theory—led to the following

¹⁸ See Landro and González (2009).

conclusions about Kolmogorov's probabilistic thinking: 1) Kolmogorov achieved a rigorous formalization of the concept of stochastic independence as a fundamental differentiating element between probability and measurement theory; 2) according to Cournot's principle and Von Mises' axiomatics, Kolmogorov assumed, in principle, a frequentist interpretation of probability; 3) given the epistemological impossibility of being able to ensure the fulfilment of the conditions of perfect equivalence and invariance of the infinite factors affecting the repetitions that constitute the foundation of the frequentist axiomatics, Kolmogorov moved towards an interpretation later known as propensionalist, on which he formulated his axiomatics; 4) consequently, the claim in the literature that Kolmogorov's axiomatics is based on a primitive idea of probability, independent of any interpretation of the nature of chance, is incorrect; 5) from the postulates of the Continuity Theorem and the consequent inclusion of the sixth axiom, provided a formal justification of the highly controversial condition of numerable additivity; 6) the impossibility of the generalization of their interpretations of the notion of probability to the weak definition of stochastic continuous processes in the time domain made the philosophical content of the Grundbegriffe less influential than its axiomatics.

REFERENCES

Bernoulli, Johann. *Ars conjectandi*. Thurnisiorum, 1713.

Bernstein, Felix. "Über eine Anwendung der Mengenlehre auf ein aus der Theorie der Säkularstörungen herrührendes Problem". *Mathematische Annalen* 71 (1912): 427-439.

<<https://doi.org/10.1007/BF02216580>>

Borel, Emile. *Éléments de la théorie des probabilités*. Hermann, 1909a.

- _____. “Les probabilités dénombrables et leurs applications arithmétiques”. *Rendiconti del Circolo Matematico di Palermo* 27 (1909b): 247-271.
<<https://doi.org/10.1007/BF03019651>>
- _____. “Sur un problème des probabilités relatif aux fractions continues”. *Oeuvres de Emile Borel*, vol. 2. Centre National de la Recherche Scientifique, 1912. 1085-1091.
- _____. *Principes et formules classiques du calcul des probabilités*. Gauthier-Villars, 1925.
- _____. “Sur les probabilités universellement négligeables”. *Comptes Rendus Hebdomadaires de Séances de l'Académie des Sciences* 190 (1930): 537-559.
- _____. *Valeur pratique et philosophie des probabilités*. Gauthier-Villars, 1939.
- Buffon, Georges-Louis. “Essai d'arithmétique morale”. *Supplément à la Histoire Naturelle* 4 (1777): 46-148.
- Cantelli, Francesco P. “La tendenza ad un limite nel censo del calcolo delle probabilità”. *Rendiconti Circolo Matematico di Palermo* 41 (1916a): 191-201.
- _____. “Sulla legge dei grandi numeri”. *Atti Reale Accademia Nazionale dei Lincei, Memorie della Classe di Scienze Fisiche, Matematiche e Naturali* 11 (1916b): 330-349.
- _____. “Sulla probabilità come limite della frequenza”. *Atti Reale Accademia Nazionale dei Lincei, Memorie della Classe di Scienze Fisiche, Matematiche e Naturali* 26 (1917a): 39-45.
- _____. “Su due applicazioni d'un teorema di G. Boole alla statistica matematica”. *Atti Reale Accademia Nazionale dei Lincei, Memorie della Classe di Scienze Fisiche, Matematiche e Naturali* 26 (1917b): 39-45.
- _____. “Una teoria astratta del calcolo delle probabilità”. *Giornale dell'Istituto degli Attuari* 3 (1932): 257-265.
<<https://doi.org/10.1122/1.2116457>>
- _____. “Consideration sur la convergence dans le calcul des probabilités”. *Annales de l'Institut Henri Poincaré* V (1935): 1-50.
- Cournot, Antoine Augustin. *Exposition de la théorie des chances et des probabilités*. Hachette, 1843.

- D'Alembert, Jean le Rond. "Doutes et questions sur le calcul des probabilités". *Mélanges de Littérature d'Histoire et de Philosophie* 5 (1767): 277-278.
- _____. "Reflexions sur le calcul des probabilités". *Opuscule mathématiques* 2 (1761): 1-25.
- Daniell, Percy J. "Integral products and probability". *American Journal of Mathematics* 43 (1921): 143-162. <<https://doi.org/10.2307/2370323>>
- Daston, Lorraine J. "D'Alembert's critique of probability theory". *Historia Mathematica* 6 (1979): 259-279. <[https://doi.org/10.1016/0315-0860\(79\)90125-3](https://doi.org/10.1016/0315-0860(79)90125-3)>
- _____. "How probabilities came to the objective and subjective". *Historia Mathematica* 21 (1994): 235-254. <<https://doi.org/10.1179/tav.1994.1994.2.235>>
- De Scheemaekere, Xavier. y Szafarz, Ariane. *The special status of mathematical probability: A storical sketch*. Centre Emile Bernheim, Research Institute in Management Sciences, 2008.
- Doob, Joseph Leo. "Kolmogorov's early work on convergence theory and foundations". *Annals of Probability* 17 (1989): 815-821.
- _____. "The development of rigor in mathematical probability, 1900-1950". *American Mathematical Monthly* 103 (1996): 586-595. <<https://doi.org/10.1080/00029890.1996.12004791>>
- Faber, Georg. "Über stetige Funktionen". *Mathematische Annalen* 69 (1910): 372-443. <<https://doi.org/10.1007/BF01456327>>
- Fréchet, Maurice. "Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements". *Rendiconti Reale Istituto Lombardo di Scienze et Lettere* 63 (1930a): 1059-1062.
- _____. "Sur la convergence 'en probabilité'". *Metron* 8 (1930b): 1-50.
- _____. *Recherches théoriques modernes sur la théorie des probabilités*. Gauthier-Villars, 1937-1938.
- Fréchet, Maurice y Halbwachs, Maurice. *Le calcul des probabilités à la portée de tous*. Dunod, 1924.
- Hadamard, Jacques. "Les principes du calcul des probabilités". *Revue de Métaphysique et de Morale* 39 (1922): 114-126.
- Hausdorff, Felix. *Grundzüge der Mengenlehre*. Von Weit, 1914.

- Hawkins, Thomas. *Lebesgue's theory of integration: Its origins and development*. University of Wisconsin Press, 1970.
- Hilbert, David. "Mathematical problems". *Bulletin of the American Mathematical Society* 8 (1902): 213-237.
- Holgate, Philip. "Independent functions: Probability and analysis in Poland between wars". *Biometrika* 84 (1997): 161-173.
- Khintchine, Aleksandr Y. "Über einen satz der Wahrscheinlichkeitsrechnung". *Fundamenta Mathematicæ* 6 (1924): 9-20.
<<https://doi.org/10.4064/fm-6-1-9-20>>
- _____. "Sur les lois de grandes nombres". *Comptes Rendus Academie des Sciences* 188 (1929): 477-480.
- _____. "Remarques sur les suites d'événements obeisant à la loi des grandes nombres". *Matematieskii Sbornik* 39 (1932a): 115-119.
- _____. "Sulle sucessioni stazionarie di evento". *Giornale dell'Istituto Italiano degli Attuari* 3 (1932b): 267-272.
- Kolmogorov, Andrei N. "The general theory of measure and the calculus of probability". *Collected works of the Mathematical Section, Communist Academy, Section for Natural and Exact Sciences* 1 (1929): 8-21.
- _____. "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung". *Mathematische Annalen* 104 (1931): 415-458.
<<https://doi.org/10.1007/BF01457949>>
- _____. *Grundbegriffe der Wabscheinlichkeitsrechnung*. Springer, 1933.
- _____. "Obituary for Evgen Evgenevich Slutsky". *Russian Mathematical Surveys* 3 (1948): 143-151. <<https://doi.org/10.1143/JPSJ.3.143>>
- _____. "On tables of random numbers". *Sankhya* 25 (1963): 369-376.
- Laemmel, Ralf. *Untersuchungen über die Ermittlung von Wahrscheinlichkeiten*. University of Zurich, 1904.
- Landro, Alberto H. y González, Mirta L. *Elementos de econometría de los fenómenos dinámicos*. Ediciones Cooperativas, 2009. <http://bibliotecadigital.econ.uba.ar/download/libros/Landro-Gonzalez_Elementos-econometria-fenomenos-dinamicos-1.pdf>.

- _____. *Acerca del problema de Bernoulli y la determinación del verdadero valor de una probabilidad*. Ediciones Cooperativas, 2016. <http://bibliotecadigital.econ.uba.ar/download/libros/Landro-Gonzalez_Acerca-del-problema-de-Bernoulli-2016.pdf>.
- Lebesgue, Henri. “Sur une généralization de l’intégrale définie”. *Comptes Rendus Académie des Sciences* 132 (1901): 1025-1028.
- _____. *Leçons sur l’intégration et la recherche des fonctions primitives*. Gauthier-Villars, 1904.
- Lévy, Paul. *Calcul des probabilités*. Gauthier-Villars, 1925.
- Łomnicki, Zbigniew. “Nouveaux fondements du calcul des probabilités (definition de la probabilité fondée sur la théorie des ensembles)”. *Fundamenta Mathematica* 4 (1923): 34-71. <<https://doi.org/10.4064/fm-4-1-34-71>>
- Łomnicki, Zbigniew y Ulam, Stanislaw. “Sur la théorie de la mesure dans les espaces combinatoires et son application au calcul des probabilités. I: Variables indépendantes”. *Fundamenta Mathematica* 23 (1934): 237-278. <<https://doi.org/10.4064/fm-23-1-237-278>>
- Nikodym, Otto. “Sur une generalization des integrals de M. J. Radon”. *Fundamenta Mathematica* 15 (1930): 131-179. <<https://doi.org/10.4064/fm-15-1-131-179>>
- Pier, Jean Paul. “Intégration et mesure 1900-1950”. *Development of mathematics 1900-1950*, Jean Paul Pier (ed.). Birkhäuser, 1994a. 517-564.
- _____. (ed.). *Development of mathematics 1900-1950*. Birkhäuser, 1994b.
- Popper, Karl R. “Probability magic or knowledge out of ignorance”. *Dialectica* 11 (1957a): 354-374.
- _____. “The propensity interpretation of probability”. *British Journal for the Philosophy of Science* 10 (1957b): 25-42.
- _____. *The logic of scientific discovery*. Hutchinson, 1959.
- Prokhorov, Yuri V. “Convergence of random processes and limit theorems of probability theory”. *Theory of Probability and its Applications* 1 (1956): 157-214.
- Radon, Johann. “Theorie und Anwendungen der absolut additiven Mengenfunktionen”. *Akademie der Wissenschaften in Wien, Sitzungsberichte, Mathematisch-naturwissenschaftliche Klasse, Abteilung* 122 (1913).

- Sierpiński, Waclaw. “Sur une définition axiomatique des ensembles mesurables”. *Bulletin International Academy of Sciences Serie A* (1918): 173-178.
- Slutsky, Yevgueni. “On the question of the logical foundation of the theory of probability”. *Bulletin of Statistics* 12 (1922): 118-119.
- _____. “Über stochastische Asymptoten und Grenzwerte”. *Metron* 5 (1925): 3-89.
- Steinhaus, Hugo. “Les probabilités dénombrables et leur rapport à la théorie de la mesure”. *Fundamenta Mathematica* 4 (1923): 286-310.
<<https://doi.org/10.4064/fm-4-1-286-310>>
- _____. “Über die Wahrscheinlichkeit dafür daß der Konvergenzkreis einer Potenzreihe ihre natürliche Grenze ist”. *Athematische Zeitschrift* 31 (1930a): 408-410.
- _____. “Sur la probabilité de la convergence de séries”. *Studia Mathematica* 2 (1930b): 21-39.
- Ulam, Stanislaw. “Zum Massbegriffe in Produkträumen”. *Verhandlung des Internationalen Mathematiker-Kongress, Zurich 2* (1932): 118-119.
- Von Kries, Johannes. *Die Principen der Wahrscheinlichkeitsrechnung. Eine logische Untersuchung*. Mohr, 1886.
- Von Mises, Richard. *Wahrscheinlichekeitsrechnung und ihre Anwendung in der statistic und teoritischen Physic*. Rosenberg, 1931.
- Von Plato, Jan. *Creating modern probability*. Cambridge University Press, 1994.
- Wiman, Anders. “Über eine Wahrcheinlichketsaufgabe bei Kettebruentwicklungen”. *Öfverseghr af Kongliga Svenska Vatenkaps-Akademiens Förhandlingar Fentindesjunde Årgangen* 57 (1900): 829-841.
- _____. *Bemerkung über eine von Glydén Aufgeworfene Wahrscheinlichkeitsfrage*. HÅkan Ohlssons Boktykeri, 1901.

EL SIMIO EGOÍSTA: CRÍTICAS A LA HIPÓTESIS DEL PENSAMIENTO MAQUIAVÉLICO*

THE SELFISH APE: CRITICS TO HYPOTHESIS OF MACHIAVELLIAN THINKING

MATEO ARIAS-VÉLEZ

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
CIUDAD DE MÉXICO, MÉXICO.
MATEOARIASVELEZ@GMAIL.COM
[HTTPS://ORCID.ORG/0000-0001-7637-4904](https://orcid.org/0000-0001-7637-4904)



RESUMEN

El objetivo de este texto es mostrar que la hipótesis del pensamiento maquiavélico de Tomasello es equivocada. Para ello, presentaré críticas a cuatro puntos fundamentales de esta hipótesis: primero, la afirmación de que los primates no-humanos no pueden atribuir falsas creencias a otros (*mindreading* o ToM) en contextos cooperativos; segundo, que estos no sienten empatía por individuos fuera de su grupo; tercero, que no tienen sentido de justicia y, por último, que no poseen algún tipo de intencionalidad conjunta. Se concluye que la evidencia científica niega que la cognición de los primates no-humanos haya evolucionado en el contexto de la competencia por la búsqueda de alimentos y que, por ello, su pensamiento sea exclusivamente competitivo e individualista. Por ende, no es posible defender la hipótesis de Tomasello.

Palabras clave: cognición animal; cooperación biológica; Tomasello; teoría de la mente; intencionalidad conjunta.

* Este artículo se debe citar: Arias-Vélez, Mateo. "El simio egoísta: críticas a la hipótesis del pensamiento maquiavélico". *Revista Colombiana de Filosofía de la Ciencia* 24.48 (2024): 137-160. <https://doi.org/10.18270/rcfc.4414>

ABSTRACT

The aim of this paper is to affirm that the Machiavellian Hypothesis of Social Thinking by Michael Tomasello is wrong. For that, I show critics to four fundamental points of this hypothesis. First, the claim that non-human primates cannot attributes false belief to other in cooperative contexts (ToM). Second, that they do not feel empathy to others out of their group. third, that they do not have fairness sense and finally, that they do not have joint intentionality. The paper concludes that scientific evidence refuses that non-human primate's cognition has evolved in a competitive context to look for food and that is why his thinking is exclusively selfish and competitive. Thus, it is not possible to defend the Tomasellos' hypothesis.

Keywords: animal cognition; biological cooperation; Tomasello; theory of mind; join intencionality.

1. INTRODUCCIÓN

Uno de los fundamentos de la hipótesis de la intencionalidad compartida de Michael Tomasello es la hipótesis del pensamiento social maquiavélico o cognición competitiva que le atribuye a todos los grandes simios con excepción del humano, quien en su opinión es esencialmente cooperativo y grupal. La hipótesis de la cognición competitiva, que retrata a los otros grandes primates como simios egoístas, está fundamentada en cuatro supuestos: 1) la capacidad social de leer la mente de otros (teoría de la mente o ToM) para cooperar, 2) la extensión de la empatía fuera de la familia, 3) el sentido de equidad y justicia, y 4) la intencionalidad conjunta son rasgos exclusivamente humanos que permitieron gestar el tipo de cognición cooperativa que poseen y que, por ende, no pueden tener los otros grandes primates.

El propósito de este texto es presentar evidencia en contra de dichos presupuestos con el ánimo de señalar que la tesis del pensamiento social maquiavélico de Tomasello es falsa. Para ello, comenzaremos con un esbozo de dicha hipótesis y sus

cuatro presupuestos, para luego lanzar críticas a cada uno (apartados del 2 al 6), y terminar concluyendo que el pensamiento primate no es esencialmente competitivo, como señala Tomasello, y por tanto la hipótesis del pensamiento maquiavélico no puede sostenerse. Además, se concluye también que, dado que todas las capacidades cognitivas que Tomasello atribuye a los humanos se comparten con los otros primates, entonces la intencionalidad conjunta es una habilidad cognitiva que ya estaba presente en el ancestro común de humanos y grandes simios.

2. COGNICIÓN COMPETITIVA O PENSAMIENTO MAQUIAVÉLICO

La hipótesis del pensamiento maquiavélico es un pilar fundamental de la hipótesis de la intencionalidad compartida e interdependencia de Michael Tomasello y le permite distinguir la cognición humana de la del resto de los primates. Tomasello afirma que la intencionalidad individual es propia de los grandes simios y sus entornos competitivos, mientras que los humanos han logrado desarrollar una forma diferente e innovadora de cognición de carácter más colaborativa: la intencionalidad compartida. Esta hipótesis propone que una serie de cambios ecológicos durante dos procesos evolutivos muy específicos forzaron a los humanos a ser cada vez más cooperativos, lo que formó en ellos mentes capaces de juntar sus intencionalidades, algo que el resto de los primates no puede hacer.

El primero de estos procesos evolutivos fue la intencionalidad conjunta. Tomasello argumenta que en ese momento el hombre pudo, a diferencia de los otros primates, extender la empatía más allá de familiares y amigos, y comenzó a formar nuevas maneras de coordinación social basadas en la atención conjunta, los compromisos conjuntos y la comunicación cooperativa entre dos socios al realizar actividades diádicas de caza y búsqueda de alimento, acciones que no pueden realizar los otros grandes primates. Todo esto permitió el surgimiento de un agente plural de tipo “nosotros”, en donde los socios que lo conformaban tenían roles determinados en la actividad basados en la igualdad y la repartición equitativa del botín (sentimiento de equidad), dando paso así a la emergencia de una moralidad de segunda

persona (Tomasello 2016 39-41). Todo ello dio paso al segundo estadio evolutivo: la intencionalidad colectiva. Aquí aparecen los grupos culturales, las actividades pasan de ser diádicas a colectivas, fundadas en un tipo de identificación moral con el grupo, que forma un autogobierno moral en el que los compromisos se ven como contratos sociales, y las ideas de bien y mal son tan colectivas que se transforman en un tipo de terreno común moral-cultural (Tomasello 2016 51).

A partir de todo esto, Tomasello afirma que la cognición humana es esencialmente cooperativa gracias a cambios ecológicos que solo vivió dicha especie, por ello, su pensamiento es único entre los primates (Tomasello 2016 124, 2019 8). En su opinión, estos últimos cuentan con una cognición basada en la competencia y la intencionalidad individual, son como simios egoístas, por lo que la búsqueda de relaciones sociales competitivas es la función propia del pensamiento primate, es decir, es para lo cual ha sido seleccionado por la naturaleza (2019 10-13). Esto último es lo que se denomina hipótesis del pensamiento maquiavélico¹ (2016 21-22), que nos dibuja a los simios no-humanos como seres egoístas en esencia. Para Tomasello este tipo de pensamiento implica que, aunque estos supuestos simios egoístas puedan cooperar con otros, sus sociedades son internamente competitivas (2022 70-76), esto es, aunque se ayuden mutuamente en algunos aspectos, compiten en los más importantes, como en la búsqueda de alimento (Tomasello 2014 20; 2016 15-23; 2019 10; 2022 70).

La propuesta básica de la hipótesis del pensamiento maquiavélico es que la inteligencia y el aumento de los cerebros en los primates es consecuencia de la competencia social entre los compañeros de grupo, porque si los competidores son los congéneres, entonces cualquier aumento en la inteligencia parece ser ventajoso y, por

² En 1976 Nicholas Humphrey propuso la hipótesis de la inteligencia social, según la cual las habilidades cognitivas de los primates habían evolucionado como una adaptación para solucionar problemas relacionados a las complejidades de la vida social (Humphrey 1976 303-317, 1983 4). Sobre esta base, se construyó, años más tarde, la hipótesis de la inteligencia maquiavélica, que apareció formalmente en el libro *Machiavellian intelligence* editado por Richard Byrne y Andrew Whiten (1988), y que ha sido la fuente sobre la que Michael Tomasello pone sus argumentos.

tanto, tal característica podría ser seleccionada y se propagaría rápido en el grupo (Byrne 1996 173).

En consecuencia, la inteligencia está relacionada con rasgos tácticos como el engaño, y la manipulación, que requieren capacidades de aprendizaje social, reconocimiento de otros y excelente memoria. Algunos autores señalaron que tales capacidades cognitivas podrían estar relacionadas con un aumento considerable del neocórtex, por lo que la hipótesis del pensamiento maquiavélico predecía que aquellas especies con habilidades para competir en contextos sociales debían desarrollar cerebros más grandes. Dado que los individuos deben ser capaces de analizar los estados mentales de otros y predecir su comportamiento, la hipótesis del pensamiento social maquiavélico está estrechamente vinculada a la llamada teoría de la mente (ToM), definida como la habilidad que tiene un individuo para atribuir estados mentales a sí mismo y a otros, y usar tales atribuciones para realizar inferencias.² Por esta última razón, también se le conoce como *mindreading* (Andrews 2020 30). Estos estados mentales pueden ser el conocimiento, la duda, las emociones, los deseos y las creencias, entre otros.

Las creencias en específico han sido fundamentales para medir el ToM en grandes primates, gracias a Daniel Dennett, quien planteó que si los chimpancés poseen una ToM similar a la de los humanos, entonces estos deberían tener creencias (y deseos) de segundo orden, esto es, creencias sobre creencias. Para poner a prueba tal suposición Dennett propuso un experimento: la prueba de la falsa creencia. La prueba tiene como fundamento un esquema del tipo “A cree que B cree que C, en donde la creencia de B es falsa, y A puede predecir el comportamiento de B a partir de esto” (Dennett 1978 569).

Tomasello fue uno de los primeros en aplicar una versión de la prueba de la falsa creencia de Dennett a simios. La primera prueba aplicada por Tomasello, junto con Call, comparaba niños y grandes primates, y señalaba que estos últimos

² La noción de teoría de la mente apareció inicialmente en un artículo de Premack y Woodruff titulado “*Does the chimpanzees have a theory of mind?*” (1978).

habían fallado al atribuir falsas creencias a otros, por lo que se concluyó que no había evidencia para apoyar la noción de que los grandes simios podían apreciar las falsas creencias de otros (Tomasello & Call 1999 381).

Con el desarrollo de nuevos métodos en la experimentación para la prueba de la falsa creencia, Tomasello refinó esta suposición. Tomasello terminó aceptando que los primates no-humanos poseían diversos estados mentales (Hare et ál. 2000 771-785; Hare et ál. 2001 139-151; Call et ál. 2004 488-498; Bräuer et ál. 2007 439-48; Tomasello 2019 4, 2022 86), incluyendo la atribución de falsas creencias, pero solo de manera implícita³ y en escenarios de competencia por alimento (Call & Tomasello 2008 191; Hare & Tomasello 2004 571-581; Tomasello 2019 72, 2022 70), dado que algunas de estas pruebas apuntaban a que los chimpancés son más hábiles en tareas competitivas que cooperativas (Krachun et ál. 2009 533-534; Krupenye et ál. 2016 113; Buttelmann et ál. 2017 8).

En resumen, Tomasello afirma que los grandes simios poseen una inteligencia social maquiavélica y competitiva, puesto que 1) no pueden atribuir falsas creencias a otros (*mindreading* o ToM) en contextos cooperativos, 2) no sienten empatía por individuos fuera de su grupo, 3) tampoco poseen un sentido de justicia y, por último, 4) no pueden llevar a cabo procesos de intencionalidad conjunta, características que sí posee la cognición humana, lo que la hace excepcional con respecto a la de los otros grandes simios. A continuación, me propongo presentar evidencia en contra de cuatro puntos fundamentales que sostienen la hipótesis del pensamiento social maquiavélico de Tomasello para mostrar que esta no es correcta.

³ Los términos 'implícito' y 'explícito' han sido usados con frecuencia en los experimentos de prueba de falsa creencia para distinguir entre la representación de un hecho (implícito) y su enunciación verbal (explícito). Para más detalles se sugiere revisar el artículo de Clements y Perner de 1994.

3. TEORÍA DE LA MENTE Y ATRIBUCIÓN DE FALSAS CREENCIAS

Tres cosas pueden señalarse de las conclusiones extraídas por Tomasello de sus experimentos sobre falsa creencia:

- a) dichos experimentos presuponen un enfoque determinista que introduce sesgos a la investigación y que no tiene en cuenta factores importantes, por lo que puede que no midan correctamente las falsas creencias. Un enfoque determinista parte de la idea de que las habilidades cognitivas y comportamentales de los animales no se ven afectadas por las condiciones de vida y de desarrollo (Boesch 2007 228). Este enfoque es característico de experimentos que toman pocos organismos de primates versus muchos humanos para concluir que los rasgos cognitivos de tales animales en cautiverio son los rasgos esenciales de la especie en general.

Boesch encuentra cinco problemas con esto: 1) los humanos con los cuales se comparan los chimpancés están en condiciones de libertad y están socialmente vinculados, mientras que los primates están en cautiverio y socialmente aislados; 2) los sujetos humanos se prueban con congéneres, mientras que los primates generalmente se prueban con individuos de otra especie; 3) los humanos interactúan directamente con los experimentadores, mientras que los primates se encuentran separados de estos; 4) en las pruebas de Tomasello los infantes humanos estaban acompañados de alguno de sus padres, mientras que los chimpancés son alejados de sus madres biológicas o familia; y 5) mientras que los sujetos humanos eran evaluados sobre tareas conspecíficas y con materiales conspecíficos, los primates eran evaluados sobre tareas y materiales humanos. La introducción de este tipo de sesgos en los experimentos es peligrosa dado que altera los resultados. Por ejemplo, Boesch señala que la mayoría de los estudios en donde se pusieron barreras a los chimpancés y a los sujetos humanos se les permitió la presencia de uno de sus padres concluyó que los humanos eran cognitivamente mejores que los chimpancés. Mientras

que en estudios que evitaron esos dos factores, los resultados resultaron ser lo contrario (Boesch 2007 234). Esto puede señalar que la presencia de los padres en los experimentos puede inducir el llamado efecto Clever Hans, en donde el sujeto estudiado recibe ayuda del experimentador u otro de manera no intencional. Por ende, Boesch cree que generalmente la introducción de este tipo de sesgos en los experimentos hace que se respondan las preguntas equivocadas.

Por ejemplo, en los experimentos de la falsa creencia de Tomasello no se estaría evaluando realmente si los chimpancés tienen una teoría de la mente, sino si tienen la capacidad de atribuirla a los humanos. Para responder la primera pregunta es necesario que se evalúen chimpancés versus chimpancés, ya que, al igual que en el caso de los humanos, el desempeño será óptimo si se evalúa su comprensión de los congéneres más que de otras especies. Por ende, Boesch señala que los enfoques comparativos son cuestionables, primero, porque no han tenido en cuenta las diferencias existentes dentro una misma especie y, segundo, porque los procedimientos experimentales se diferencian sistemáticamente de una especie a otra.

En síntesis, las diferencias ecológicas de crianza son fundamentales en el desarrollo de habilidades cognitivas y, por tanto, deben ser tenidas en cuenta en los experimentos de comparación de especies. Este tipo de estudios debe también integrar tanto la etología y la psicología, como la ecología y los estudios de desarrollo (Boesch 2008 454).⁴ En

4 Tomasello y Call respondieron a las críticas de Boesch diciendo que muchas de las variables que Boesch señala no pueden ser eliminadas de los experimentos, pero en su opinión sí pueden ser controladas (Tomasello & Call 2007 450). Boesch señaló que esto no es del todo cierto, y menos en los experimentos realizados por Tomasello, señalando, además, que en los videos que acompañan muchas de sus pruebas puede apreciarse el efecto Clever Hans. Por ejemplo, no hay descripciones de Tomasello y Call en sus pruebas en donde se realice algún tipo de control que minimice la diferencia en el desempeño que pueden tener niños y crías primates en ausencia o presencia de sus

coherencia con lo anterior, en un reciente artículo de Ghrear et ál. se pone a prueba el test de Sally-Anne, que ha sido el paradigma para el diseño de experimentos sobre creencias falsas con niños y primates. Se encontró que en los diseños clásicos, como los de Tomasello, había un sesgo de conocimiento en los niños que interfería en la prueba, por lo que se concluye que los experimentos clásicos no median realmente la comprensión de creencias falsas, como Boesch había señalado en el 2007. Así, es probable que la prueba de Tomasello y Call de 1999 no estuviera midiendo realmente las habilidades cognitivas para creencias falsas ni en los niños ni en los primates.

- b) Por otra parte, aunque aceptáramos que los experimentos de Tomasello no están sesgados y miden con precisión la falsa creencia y la capacidad de *mindreading*, podríamos notar que sus conclusiones son algo extrañas y parecen apuntar a una tesis contraria; los grandes primates pueden atribuir correctamente falsas creencias en contextos cooperativos (Buttelmann et ál. 2017 1-13), lo cual sugieren tanto el artículo de Krupenye et ál. (2016) y con mayor fuerza el de Buttelmann et ál. No obstante, Tomasello continuaría insistiendo en que los grandes primates son esencialmente competitivos o que solo pueden tener una teoría de la mente en contextos competitivos (Tomasello 2019 72, 2022 86), aunque desde el 2017 su laboratorio ya hubiera probado que los chimpancés pueden leer la mente de otros en escenarios de ayuda.

padres (2007 454). A pesar del acervo de evidencia presentada por Boesch en el 2007, Tomasello y Call señalan que no hay evidencia de que existan diferencias cognitivas entre primates en cautiverio y en estado de libertad, y que tal regla se aplica igualmente a los niños occidentales con respecto al resto de niños de todo el mundo (2007 451). Por otra parte, Tomasello y Call señalan que las pruebas a las que se exponen los chimpancés en cautiverio los ponen en situaciones más desafiantes que las que podrían encontrar en la vida salvaje, lo que expondría mejor sus habilidades cognitivas, no obstante, este contraargumento de Tomasello y Call parece más una suposición, pues no existe material de campo con el cual compararse (2007 450).

- (c) Si aceptamos lo anterior, entonces la hipótesis del pensamiento maquiavélico puede no ser adecuada y se necesitan otro tipo de propuestas. Una alternativa más coherente puede ser la de Kristen Andrews, quien señala que la hipótesis del pensamiento maquiavélico no ha explicado realmente el surgimiento del pensamiento, pues no es necesario atribuirle estados mentales a otro para engañarlo (Andrews 2020 186). Esto la lleva a proponer, basada en el trabajo de Allison Jolly, la hipótesis del pensamiento menciano, que señala que el pensamiento es producto de tratar de resolver problemas sociales a través de la cooperación: Andrews señala que la capacidad de pensar en los deseos, creencias e intenciones de otros pudo evolucionar para promover interacciones sociales colaborativas (Andrews 2018 91, 2020 186). En este sentido, Andrews sugiere que la función de una teoría de la mente (mindreading) no es predecir el comportamiento de los otros, sino explicar (o justificar) tal comportamiento, es decir, los primates adjudican creencias a otros para dar cuenta de acciones particulares, en especial cuando el comportamiento es raro o va en contra de lo normalmente aceptado (2018 100), porque un agente debe poder explicar y comprender el comportamiento de otra persona antes de poder predecir qué hará. Lo anterior permite, en teoría, reducir la disonancia cognitiva ocasionada por el comportamiento inusual del otro y conservar la relación cooperativa de ambos. Así, los primates no buscan leer las creencias y deseos de los otros para engañarlos, sino para comprenderlos y poder trabajar con ellos en tareas conjuntas, lo que garantiza una cohesión social que es fundamental para el establecimiento y sostenimiento de cualquier tipo de sociedad. En síntesis, la ToM, en este contexto puede tener la función de predecir el comportamiento de los otros no para engañarlos, sino para ayudarlos.

Dada la alta cantidad de evidencia empírica que soporta la idea de que los grandes primates, y especialmente los humanos, son seres cooperadores, tendría más sentido pensar que su inteligencia social se ha desarrollado en escenarios cooperativos y de ayuda mutua que

lo contrario, esto es, hay mayor evidencia que soporte la hipótesis del pensamiento social menciano que maquiavélico.⁵ En otras palabras, el supuesto 1 de Tomasello es equivocado y la evidencia empírica parece señalar lo contrario: los primates no-humanos no son, en términos cognitivos, esencialmente competitivos.

4. LA EMPATÍA MÁS ALLÁ DE FAMILIA Y AMIGOS

Para Tomasello, dado que los grandes primates poseen un pensamiento individual y competitivo, entonces son incapaces de extender la empatía más allá de familiares y amigos, algo que solo lograron los humanos (supuesto 2). Tomasello (2016 49, 2019 223) argumenta que los grandes primates solo ayudan si su acción no es costosa y puede ser retribuida, algo que con familiares y amigos es más probable (reciprocidad). Pretendo criticar esta suposición basándome en el trabajo de Frans de Waal.

Desde comienzos de la década del 2000, Frans de Waal propuso que la empatía era un rasgo que también aparecía en otras especies, dado que puede observarse cómo diferentes organismos se ven afectados emocionalmente por otros (Preston & De Waal 2002a 284). En este orden de ideas, De Waal propone que la empatía es un fenómeno que se presenta en diversos niveles de complejidad y que está fundamentada en mecanismos próximos (modelo de percepción-acción) y últimos que pueden ser estudiados empíricamente (de Waal & Preston 2017). Para explicar su funcionamiento en términos evolutivos, De Waal presenta un modelo para la empatía compuesto por tres niveles evolutivos: el contagio emocional, entendido como un tipo de vínculo fisiológico, la preocupación comprensiva (manifiesta en acciones

⁵ Aunque hay mucha evidencia que soporta la tesis del pensamiento social de Humprey, lo que incluso ha llevado a consolidarlo como uno de los paradigmas de la psicología evolutiva y comparada, no hay material empírico que sustente la hipótesis del pensamiento maquiavélico, a excepción de un modelo matemático para el desarrollo de la inteligencia publicado por Gavrilets y Vose (2006), sin embargo, este es solo una propuesta hipotética, carente de sustento empírico.

como la compasión y el consuelo), y, por último, la toma de perspectiva empática, esto es, la capacidad de adoptar el punto de vista del otro (de Waal 2008 282-286).

Tanto el trabajo experimental de De Waal como las observaciones de campo le han llevado a señalar que los grandes primates pueden establecer relaciones afectivas basadas en la empatía no solo con organismos que son sus socios o familiares, sino incluso con individuos de otras especies diferentes a la suya. Para probar experimentalmente lo anterior, De Waal midió bostezos contagiosos en chimpancés, demostrando que los chimpancés se contagian de los bostezos cuando los veían expresarse en humanos y congéneres (Campbell y De Waal 2014 3).

Igualmente, de Waal ha señalado en artículos y libros que las motivaciones empáticas en primates y humanos no se basan en la espera de algún tipo de compensación o reciprocidad, sino que los chimpancés y bonobos son sumamente emocionales y cooperativos, y tienen tanta disposición de ayudar a otros como cualquier humano (de Waal 2022 134-135). A favor de lo anterior, De Waal ha presentado suma evidencia observacional de acciones empáticas, como el consuelo en bonobos (de Waal 2022 117), o incluso de relaciones de amistad entre bonobos y humanos (de Waal 2019 28) que, como bien señala Scotto (2015 430), por su semejanza con los vínculos emocionales que establecemos los humanos, pueden ser llamados propiamente amistades.

Además, existe evidencia más reciente de empatía fuera de familiares y amigos en observaciones de campo con bonobos en las selvas del Congo, donde se han registrado asociaciones extragrupales de hembras no emparentadas (Tokuyama & Furuichi 2016 27-35), y muestras de ayuda y empatía fuera de su grupo (Tokuyama et ál. 2021 1-11), como fue el caso de la adopción de huérfanos provenientes de otro grupo, lo que garantiza que no había parentesco ni ningún tipo de cercanía social.

En conclusión, existe evidencia científica, tanto experimental como observacional, que demuestra que los grandes primates son capaces de extender la empatía más allá de familiares y amigos, una capacidad que Tomasello atribuye exclusivamente a los seres humanos. Toda esta evidencia a favor de la empatía extrafamiliar en grandes simios está altamente registrada en grupos de chimpancés y bonobos en libertad y cautiverio. Lo común que es la extensión del sentimiento de empatía en

chimpancés, bonobos y humanos no solo contradice el segundo supuesto del pensamiento maquiavélico de Tomasello, sino que también nos sugiere que ya ocurría antes en la filogenia primate y, por ende, no es un comportamiento que haya surgido inicialmente en los humanos actuales.

5. EQUIDAD Y JUSTICIA EN PRIMATES

El tercer presupuesto que Tomasello deriva de la hipótesis de la cognición competitiva y que le sirve para señalar que los grandes primates son esencialmente competitivos mientras que los humanos son súpercooperadores es que, según él, establecer relaciones no solo de empatía sino de equidad (justicia) es algo que no pueden hacer los otros grandes simios, sino que es un rasgo exclusivo de la cognición humana (Tomasello 2016 32). Similar al argumento anterior, sugiero que en la actualidad existe evidencia que niega tal aseveración porque con los trabajos de Sarah Brosnan se muestra que los chimpancés, e incluso ciertos monos, ya tienen un sentido de equidad o justicia entre ellos.

Experimentos a principios de los noventa señalaron que el comportamiento de los chimpancés y otros primates parece reflejar un sentido de regularidad social, que puede ser un precursor del sentido de justicia (De Waal 1991 335-349). Brosnan y De Waal (2003 297) demostraron que los monos capuchinos marrones (*Cebus apella*) se negaban a participar de un experimento si veían que un congénere suyo obtenía una recompensa más atractiva (uvas en vez de pepino) por el mismo esfuerzo. Los autores llaman a esto sentido de aversión a la desigualdad distributiva, que puede manifestarse a través de respuestas a la inequidad, como el rechazo a participar de una tarea (Brosnan 2006 173). Tomasello ha sido crítico de este experimento sosteniendo que no tuvo las condiciones de control necesarias (2016 166).⁶ No obstante, Brosnan continuó con sus investigaciones aplicándolas a otros

⁶ Señalando que los experimentos de Brosnan no contaban con las condiciones de control necesarias, en el 2006 Bräuer, Call y Tomasello buscaron replicar el experimento anterior, concluyendo que

primates: los chimpancés. Esto la llevó a concluir que también muestran una respuesta a la inequidad de las recompensas, la cual incluso puede variar de un grupo a otro (Brosnan, Shiff & de Wall 2005 256-257). Posteriormente, Brosnan, Freeman y De Waal hallaron que en contextos cooperativos los capuchinos dejaban de cooperar con compañeros que dominaban las mejores recompensas (2006 713-724).

En el 2010, Brosnan estudió nuevamente las respuestas de igualdad en chimpancés con el ánimo de revisar si había afectación en sus respuestas aversivas a la desigualdad según factores sociales o según el esfuerzo y el tipo de tarea. Brosnan encontró que no hay aversión a la desigualdad si las recompensas son gratis, esto es, se requiere una tarea para que haya una respuesta a la desigualdad (Brosnan 2010b et ál. 1233). Además, no hubo cambios en el comportamiento según el tipo de tarea o esfuerzo, solo si había una distribución desigual. Lo interesante de esto es que Brosnan señaló que también había aversión a la desigualdad por parte del sujeto que recibía la mejor recompensa y no solo del afectado, un tipo de respuesta de la que no se tenía registro antes (Brosnan et ál. 2010b 1235). También concluyeron que existe una interacción entre las expectativas sociales e individuales que dependen fundamentalmente de factores sociales, esto es, del rol de los hombres o las mujeres en el grupo. Estos resultados se confirmaron luego de analizar la relación entre preferencias prosociales e inequidad, encontrando que los monos capuchinos marrones preferían no recibir una recompensa si sabían que su compañero no obtendría nada (Brosnan et ál. 2010a, 279-288). Así, la respuesta a la desigualdad puede ser negarse a participar en una tarea si la recompensa es desventajosa para sí mismo o para el compañero.

la respuesta de los chimpancés era contraria al patrón de rechazo señalado por Brosnan y que, por el contrario, cuando un chimpancé veía que otro obtenía una mejor recompensa, estos suplicaban al experimentador para que les diera mejor comida, esto es, no sentían aversión a la desigualdad, sino que desarrollaban una expectativa de que ellos también recibirían mejor comida (Bräuer, Call & Tomasello 2006 3126). Sin embargo, los experimentos de Brosnan tanto en monos como en simios sugieren con alto respaldo experimental que la aversión a la inequidad existe y está bastante extendida entre los primates.

Revisando estos hallazgos, Brosnan propuso en el 2011 una explicación: la hipótesis de la cooperación, según la cual la aversión a la inequidad es un mecanismo para promover relaciones cooperativas exitosas a largo plazo entre individuos que no son parientes, dado que permite reconocer y evitar a socios que luego de completar una tarea cooperativa toman más de lo que les corresponde (Brosnan 2011). Así, este comportamiento evolucionó junto con la cooperación y puede representar una adaptación para aumentar los beneficios relacionados con interacciones cooperativas en primates no humanos (Brosnan 2023 245). Por ello, Brosnan y De Waal proponen que la evolución de una respuesta aversiva a la desigualdad puede ser un precursor del sentido de justicia humana (Brosnan & De Waal 2014 314), y que estudiarla en otras especies de primates puede arrojar luces sobre su desarrollo y función (Brosnan 2013 268).

En síntesis, puede afirmarse que existe amplia y reciente evidencia que señala que los humanos no son la única especie que responde negativamente a la inequidad y que, por el contrario, parece un rasgo adaptativo que puede relacionarse con la prosocialidad y la cooperación en primates (Brosnan & de Waal 2009 285-301). Este cuerpo de evidencia experimental contradice el tercer supuesto de la hipótesis de la inteligencia maquiavélica de Tomasello, según el cual los humanos son la única especie con un sentido de equidad y los otros primates carecen de ello por ser esencialmente competitivos (Tomasello 2016 32).

6. LA INTENCIONALIDAD CONJUNTA NO ES EXCLUSIVAMENTE HUMANA

El último supuesto (4) señala que, dada la esencial cognición primate para competir en la búsqueda de alimento, los simios cazan solos y únicamente los humanos tienen una intencionalidad conjunta, que comprende habilidades como la elección y el control de socios en actividades conjuntas o colaborativas, la atención conjunta y la comunicación cooperativa (Tomasello 2019 57). Sin embargo, existe fuerte evidencia de lo contrario. Schino y Aureli (2009 59, 2017 671) han detallado la impor-

tancia de la elección y control de socios en primates, quienes establecen relaciones cooperativas basadas en el altruismo recíproco, y donde las emociones funcionan como mecanismo importante de elección.

En este sentido, y contrario a lo que Tomasello señala, los primates no humanos son sumamente cooperativos, cazan y comparten alimentos con socios de su grupo. La evidencia más reciente apunta a que incluso simios como los monos capuchinos pueden cooperar con más de un individuo a la hora de obtener y compartir comida, y que, además, la elección del socio para esta actividad está basada en beneficios para los compañeros a largo plazo (Sabbatini et ál. 2012 579). Por ejemplo, con respecto a la atención conjunta, Tomasello (2014 47, 2019 57) argumenta que los grandes simios entienden que sus congéneres ven cosas y pueden mirar en la misma dirección, pero no interactúan con los otros cuando de actividades de atención conjunta se trata. Contrariamente, Leavens y Racine (2009) encontraron que los grandes simios muestran todas las características comunes de la atención conjunta, similar a la presentada por los humanos (257), lo que indica que, diferente a lo señalado por Tomasello, los humanos no son los únicos que pueden orientar su atención a metas comunes (Bader 2018 157-158).

Para Tomasello (2014, 2016, 2019), otra de las características de una intencionalidad conjunta es la idea de que los socios pueden generar compromisos conjuntos, habilidades específicas de los humanos y que debieron aparecer con los neandertales hace cuatrocientos mil millones de años. No obstante, existe evidencia de compromisos conjuntos en bonobos y otros primates. Por ejemplo, Heesen et ál. hicieron experimentos con bonobos y chimpancés, y concluyeron que, al igual que los humanos, estos se involucran en actividades de compromisos conjuntos basadas en un sentido de obligación mutua, con roles interactivos y comunicación gestual, además de que son conocedores de las consecuencias de romper tales compromisos, por lo que luego de que se les interrumpa en una actividad se les ve bastante dispuestos a reanudarla con su compañero (Heesen et ál. 2020 4-7, 2021b 13; Bangertner et ál 2022 6-7).

Además, los bonobos han mostrado ser capaces de comunicarse gestualmente, intercambiando miradas y señales antes y después de comprometerse en actividades

de compromiso conjunto, algo similar a lo que hacen los humanos. Esto ha llevado a algunos investigadores a señalar que el compromiso conjunto no es un proceso colaborativo único de la especie humana ni surgió con el género *Homo*, sino que, por el contrario a lo señalado por Tomasello, ya se encontraba presente en el último ancestro común con el género *Pan* (Heesen et ál. 2021a 10).

Así pues, si los grandes primates cuentan con atención conjunta, compromiso conjunto y comunicación cooperativa, entonces no hay razones para sostener que no poseen una intencionalidad conjunta, habilidad cognitiva que Tomasello cree que surgió únicamente en el linaje humano y que ha sido de importancia para que la especie tenga habilidades cognitivas y cooperativas que en su opinión le son exclusivas (2016 124, 2019 8). Una prueba de que la intencionalidad conjunta no es propiamente humana puede encontrarse en el análisis de juegos sociales triádicos en bonobos, donde se demostró que estos se comportaban de manera similar a como lo hacen los niños humanos en las mismas actividades, mostrando interés en la actividad conjunta más que en los objetos, y usando gestos comunicativos para alentar a los socios a cumplir su rol (Pika & Zuberbühler 2008 209-210). Por ende, y teniendo en cuenta estas evidencias empíricas, podemos señalar en contra de Tomasello que la intencionalidad conjunta no es exclusivamente humana y debe haber surgido mucho antes que el linaje homínido.

7. CONCLUSIÓN

Hemos presentado evidencia empírica desarrollada en las últimas décadas en contra de cuatro supuestos básicos de la hipótesis del pensamiento social maquiavélico sostenida por Michael Tomasello, mostrando que no hay razones para sostener que los primates no humanos sean simios egoístas incapaces de atribuir falsas creencias a otros en contextos cooperativos y, por ende, que no tengan una ToM; segundo, que estos no sientan empatía por individuos fuera de su grupo; tercero, que no tengan sentido de justicia y, por último, que no posean una intencionalidad conjunta. En tal caso, puede señalarse que la hipótesis del pensamiento maquiavélico o cognición

competitiva es una propuesta equivocada para explicar la cognición animal. Además, dado que la hipótesis del pensamiento maquiavélico es el fundamento de la propuesta de la intencionalidad compartida, entonces esta última también está equivocada, y puede concluirse que no es diferenciadora de la cognición humana y animal o que brinde algún tipo de excepcionalidad a la primera.

Así, se concluye que la evidencia científica contradice la hipótesis de que cognición de los primates no humanos haya evolucionado en el contexto de la competencia por la búsqueda de alimentos y que, por ello, su pensamiento sea exclusivamente competitivo e individualista. Por lo tanto, no es posible defender la hipótesis de Tomasello, sino que, por el contrario, hay material empírico suficiente para mostrar que los grandes primates no son simios egoístas, sino seres cooperativos que poseen procesos cognitivos similares a los humanos.

Como consecuencia de lo anterior, se afirma que todas las características que Tomasello cree que son únicas del *Sapiens* ya se encontraban en otros homínidos o incluso en otros primates. Igualmente, puede concluirse que si existe evidencia de ToM, empatía, sentido de justicia e intencionalidad conjunta en primates, podemos deducir que estas habilidades sociocognitivas ya estaban presentes en el ancestro común del género *Homo* y *Pan*. Por ende, debe ser posible encontrar aspectos básicos de la intencionalidad colectiva, que Tomasello relaciona con la cultura, en otros primates, como los chimpancés y bonobos, lo que implicaría repensar lo que sabemos sobre los humanos y su desarrollo evolutivo como especie.

REFERENCIAS

- Andrews, Kristen. "Do Apes Attribute Beliefs to Predict Behavior?: A Mengzian Social Intelligence Hypothesis". *The Harvard Review of Philosophy* 25.1 (2018): 89-110. <<https://doi.org/10.5840/harvardreview201892117>>
- _____. *The Animal Mind: An Introduction to the Philosophy of Animal Cognition*. Londres: Routledge, 2020. <<https://doi.org/10.4324/9780203712511>>

- Bader, Oren. “Being in a Group—the Emotional Scaffolding of the Evolution of Collective Intentionality”. *Landscapes of Collectivity in the Life Sciences*. Eds. Snait B. Gissis, Ehud Lamm y Ayelet Shavit. Londres: The MIT Press, 2018. 155-168.
- Bangerter, Adrian., et ál. “Every Product Needs a Process: Unpacking Joint Commitment as a Process Across Species”. *Philosophical Transactions of the Royal Society of London* 377.1859 (2022): Online. <<https://doi.org/10.1098/rstb.2021.0095>>
- Boesch, Christophe. “What Makes Us Human (Homo Sapiens)? The Challenge of Cognitive Cross-Species Comparison”. *Journal of Comparative Psychology* 121.3 (2007): 227-240. <<https://psycnet.apa.org/doi/10.1037/0735-7036.121.3.227>>
- _____. “Taking Development and Ecology Seriously When Comparing Cognition: Reply to Tomasello and Call (2008)”. *Journal of Comparative Psychology* 122.4 (2008): 453-455. <<https://doi.org/10.1037/0735-7036.122.4.453>>
- Bräuer, Juliane., Call, Josep., y Tomasello, Michael. “Are Apes Really Inequity Averse?”. *Proceedings Biological Sciences* 273.1605 (2006): 3123-3128. <<https://doi.org/10.1098/rspb.2006.3693>>
- _____. “Chimpanzees Really Know What Others Can See in a Competitive Situation”. *Animal Cognition* 10.4 (2007): 439-448. <<https://doi.org/10.1007/s10071-007-0088-1>>
- Brosnan, Sarah F. “Nonhuman Species Reactions to Inequity and Their Implications for Fairness”. *Social Justice Research* 19.1 (2006): 153-185. <<https://doi.org/10.1007/PL00022136>>
- _____. “Responses to Inequity in Non-human Primates”. *Neuroeconomics: Decision making and the brain*. Eds. Paul W. Glimcher, Colin F. Camerer, Ernst Fehr y Russell A. Poldrack. Londres: Elsevier Academic Press, 2009. 285-301. <doi.org/10.1016/B978-0-12-374176-9.00019-1>.
- _____. “A Hypothesis of the Co-evolution of Cooperation and Responses to Inequity”. *Frontiers in Neuroscience* 5.43 (2011): Online. <<https://doi.org/10.3389/fnins.2011.00043>>

- _____. “Comportamientos relacionados con la justicia y la equidad en primates no humanos”. *Ludus Vitalis* 21.40 (2013): 249-272. <https://www.ludus-vitalis.org/html/textos/40/40-13_brosnan.pdf>
- _____. “A Comparative Perspective on the Human Sense of Justice”. *Evolution and Human Behavior* 44.3 (2023): 242-249. <<https://doi.org/10.1016/j.evolhumbehav.2022.12.002>>
- Brosnan, Sarah F., Schiff, Hillary C., y de Waal, Frans B. M. “Tolerance for Inequity May Increase with Social Closeness in Chimpanzees”. *Proceedings Biological Sciences* 272.1560 (2005): 253-258. <<https://doi.org/10.1098/rspb.2004.2947>>
- Brosnan, Sarah F., Freeman, Cassiopeia., y de Waal, Frans B.M. “Partner’s Behavior, not Reward Distribution, Determines Success in an Unequal Cooperative Task in Capuchin Monkeys”. *American Journal of Primatology* 68.7 (2006): 713-724. <<https://doi.org/10.1002/ajp.20261>>
- Brosnan, Sarah F., et ál. “Competing Demands of Prosociality and Equity in Monkeys”. *Evolution and Human Behavior* 31.4 (2010a): 279-288. <<https://doi.org/10.1016/j.evolhumbehav.2010.02.003>>
- _____. “Mechanisms Underlying Responses to Inequitable Outcomes in Chimpanzees, Pan Troglodytes”. *Animal Behaviour* 79.6 (2010b): 1229-1237. <<https://doi.org/10.1016/j.anbehav.2010.02.019>>
- Brosnan, Sarah F y Frans B. de Waal. “Monkeys Reject Unequal Pay”. *Nature* 425.6955 (2003): 297-299. <<https://doi.org/10.1038/nature01963>>
- _____. “Cebus apella Tolerate Intermittent Unreliability in Human Experimenters”. *International Journal of Primatology* 30.5 (2009): 663-674. <<https://doi.org/10.1007/s10764-009-9366-x>>
- _____. “Evolution of Responses to (un)Fairness”. *Science* 346.6207 (2014): 1251776. <<https://doi.org/10.1126/science.1251776>>
- Buttelmann, David et ál. “Great Apes Distinguish True from False Beliefs in an Interactive Helping Task”. *PloS one* 12.4 (2017): online. <<https://doi.org/10.1371/journal.pone.0173793>>

- Byrne, Richard W. “Machiavellian Intelligence”. *Evolutionary Anthropology* 5.5 (1996): 172-180. <[https://doi.org/10.1002/\(SICI\)1520-6505\(1996\)5:5%3C172::AID-EVAN6%3E3.0.CO;2-H](https://doi.org/10.1002/(SICI)1520-6505(1996)5:5%3C172::AID-EVAN6%3E3.0.CO;2-H)>
- Byrne, Richard W. y Andrew Whiten (Eds.). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans*. Oxford University Press, 1988. <<https://api.semanticscholar.org/CorpusID:140446133>>
- Call, Josep., et ál. “‘Unwilling’ versus ‘unable’: Chimpanzees’ Understanding of Human Intentional Action”. *Developmental Science* 7.4 (2004): 488-498. <<https://doi.org/10.1111/j.1467-7687.2004.00368.x>>
- Call, Josep y Michael Tomasello. “Does the Chimpanzee Have a Theory of Mind? 30 years Later”. *Trends in Cognitive Sciences* 12.5 (2008): 187-192. <<https://doi.org/10.1016/j.tics.2008.02.010>>
- Campbell, Matthew W. y Frans B. M. de Waal. “Chimpanzees Empathize with Group Mates and Humans, But not with Baboons or Unfamiliar Chimpanzees”. *Proceedings Biological Sciences* 281.1782 (2014): Online. <<https://doi.org/10.1098/rspb.2014.0013>>
- Clements, Wendy A. y Josef Perner. “Implicit Understanding of Belief”. *Cognitive Development* 9.4 (1994): 377-395. <[https://doi.org/10.1016/0885-2014\(94\)90012-4](https://doi.org/10.1016/0885-2014(94)90012-4)>
- de Waal, Frans B. M. “The Chimpanzee’s Sense of Social Regularity and its Relation to the Human Sense of Justice”. *American Behavioral Scientist* 34.3 (1991): 335-349. <<https://doi.org/10.1177/0002764291034003005>>
- _____. “Putting the Altruism Back into Altruism: The Evolution of Empathy”. *Annual Review of Psychology* 59.1 (2008): 279-300. <<https://doi.org/10.1146/annurev.psych.59.103006.093625>>
- _____. *El último abrazo. Las emociones animales y lo que nos cuentan de nosotros*. Barcelona: Tusquets, 2019.
- _____. *La edad de la empatía. ¿Somos altruistas por naturaleza?* Barcelona: Tusquets editores, 2022.

- de Waal, Frans B. M. y Stephanie D. Preston. "Mammalian Empathy: Behavioural Manifestations and neural basis". *Nature Reviews. Neuroscience* 18.8 (2017): 498-509. <<https://doi.org/10.1038/nrn.2017.72>>
- Dennett, Daniel. "Beliefs about Beliefs". *Behavioral and Brain Sciences* 1.4 (1978): 568-570. <<https://doi.org/10.1017/S0140525X00076664>>
- Gavrilets, Sergey y Aaron Vose. "The Dynamics of Machiavellian Intelligence". *Proceedings of the National Academy of Sciences of the United States of America* 103.45 (2006): 16823-16828. <<https://doi.org/10.1073/pnas.0601428103>>
- Ghrear, Siba *et ál.* "Are the Classic False Belief Tasks Cursed? Young Children Are Just as Likely as Older Children to Pass a False Belief Task When They Are Not Required to Overcome the Curse of Knowledge". *PLoS one* 16.2 (2021): <<https://doi.org/10.1371/journal.pone.0244141>>
- Hare, Brian *et ál.* "Chimpanzees Know What Conspecifics Do and Do Not See". *Animal Behaviour* 59.4 (2000): 771-785. <<https://doi.org/10.1006/anbe.1999.1377>>
- _____. "Do Chimpanzees Know What Conspecifics Know?" *Animal Behaviour* 61.1 (2001): 139-151. <<https://doi.org/10.1006/anbe.2000.1518>>
- Hare, Brian., y Michael Tomasello. "Chimpanzees are More Skilful in Competitive than in Cooperative Tasks". *Animal Behaviour* 68.1 (2004): 571-581. <<https://doi.org/10.1016/j.anbehav.2003.11.011>>
- Heesen, Raphaela *et ál.* "Bonobos Engage in Joint Commitment". *Science Advances* 6.51 (2020): <<https://doi.org/10.1126/sciadv.abd1306>>
- _____. "Assessing Joint Commitment as a Process in Great Apes". *iScience* 24.8 (2021a): <<https://doi.org/10.1016/j.isci.2021.102872>>
- _____. "Evidence of Joint Commitment in Great Apes' Natural Joint Actions". *Royal Society Open Science* 8.12 (2021b): *Online*. <<https://doi.org/10.1098/rsos.211121>>
- Humphrey, Nicholas K. "The Social Function of Intellect". *Growing Points in Ethology*. Eds. Patrick P. G. Bateson y Robert A. Hinde. Cambridge University Press, 1976. 303-317. <<https://api.semanticscholar.org/CorpusID:17845758>>

- _____. *Consciousness Regained: Chapters in the Development of Mind*. Oxford University Press, 1983.
- Krachun, Carla et ál. “A Competitive Nonverbal False Belief Task for Children and Apes”. *Developmental Science* 12.4 (2009): 521-535. <<https://doi.org/10.1111/j.1467-7687.2008.00793.x>>
- Krupenye, Christopher et ál. “Great Apes Anticipate that Other Individuals Will Act According to False Beliefs”. *Science* 354.6308 (2016): 110-114. <<https://doi.org/10.1126/science.aaf8110>>
- Leavens, David A. y Timothy P. Racine. “Joint Attention in Apes and Humans: Are humans Unique?”. *Journal of Consciousness Studies* 16.6-8 (2009): 240-267. <<https://www.ingentaconnect.com/contentone/imp/jcs/2009/00000016/f0030006/art00010>>
- Pika, Simone y Klaus Zuberbühler. “Social Games Between Bonobos and Humans: Evidence for Shared Intentionality?”. *American Journal of Primatology* 70.3 (2008): 207-210. <<https://doi.org/10.1002/ajp.20469>>
- Premack, David y Guy Woodruff. “Does the Chimpanzee Have a Theory of Mind?”. *Behavioral and Brain Sciences* 1.4 (1978): 515-526. <<https://doi.org/10.1017/S0140525X00076512>>
- Preston, Stephanie D. y Frans B. M. de Waal. “Empathy: Its Ultimate and Proximate Bases”. *The Behavioral and Brain Sciences* 25.1 (2002a): 1-20. <<https://doi.org/10.1017/s0140525x02000018>>
- _____. “The Communication of Emotions and the Possibility of Empathy in Animals”. *Altruism and Altruistic Love: Science, Philosophy, and Religion in Dialogue*. Eds. Stephn G. Post, Lynn G. Underwood, Jeffrey P. Schloss y William B. Hurlbut. Oxford University Press, 2002b. 284-308. <<https://psycnet.apa.org/doi/10.1093/acprof:oso/9780195143584.003.0025>>
- Sabbatini, Gloria., et ál. “Food Transfers in Capuchin Monkeys: An Experiment on Partner Choice”. *Biology Letters* 8.5 (2012): 757-759. <<https://doi.org/10.1098/rsbl.2012.0534>>

- Schino, Gabriele y Filippo Aureli. "Reciprocal Altruism in Primates: Partner Choice, Cognition, and Emotions". *Advances in the Study of Behavior* 39.1 (2009): 45-69. <[https://doi.org/10.1016/S0065-3454\(09\)39002-6](https://doi.org/10.1016/S0065-3454(09)39002-6)>
- _____. "Reciprocity in Group-living Animals: Partner Control Versus Partner Choice". *Biological Reviews of the Cambridge Philosophical Society* 92.2 (2017): 665-672. <<https://doi.org/10.1111/brv.12248>>
- Scotto, Carolina. "Empatía, antropomorfismo y cognición animal". *Principia: An International Journal of Epistemology* 19.3 (2015): 423-452. <<https://doi.org/10.5007/1808-1711.2015V19N3P423>>
- Tokuyama, Nahoko y Takeshi Furuichi. "Do Friends Help Each Other? Patterns of Female Coalition Formation in Wild Bonobos at Wamba". *Animal Behaviour* 119.1 (2016): 27-35. <<https://doi.org/10.1016/j.anbehav.2016.06.021>>
- Tokuyama, Nahoko *et ál.* "Two Wild Female Bonobos Adopted Infants from a Different Social Group at Wamba". *Scientific Reports* 11.1 (2021): <<https://doi.org/10.1038/s41598-021-83667-2>>
- Tomasello, Michael. *A Natural History of Human Thinking*. Harvard University Press, 2014.
- _____. *A Natural History of Human Morality*. Harvard University Press, 2016.
- _____. *Becoming Human: A Theory of Ontogeny*. Belknap Press of Harvard University Press, 2019.
- _____. *The Evolution of Agency: Behavioral Organization from Lizards to Humans*. Cambridge: The MIT Press, 2022.
- Tomasello, Michael, y Josep Call. "A Nonverbal False Belief Task: The Performance of Children and Great Apes". *Child Development* 70.2 (1999): 381-395. <<https://doi.org/10.1111/1467-8624.00028>>
- _____. "Assessing the Validity of Ape-human Comparisons: a Reply to Boesch (2007)". *Journal of Comparative Psychology* 122.4 (2008): 449-52. <<https://doi.org/10.1037/0735-7036.122.4.453>>

LA POSIBILIDAD DE EXPLICACIÓN CIENTÍFICA A PARTIR DE MODELOS BASADOS EN REDES NEURONALES ARTIFICIALES*

THE POSSIBILITY OF SCIENTIFIC EXPLANATION FROM MODELS BASED ON ARTIFICIAL NEURAL NETWORKS

ALEJANDRO E. RODRÍGUEZ-SÁNCHEZ
Universidad Panamericana, Facultad de Ingeniería
Zapopan, México.
aerodriguez@up.edu.mx
<https://orcid.org/0000-0003-3397-5261>



RESUMEN

En inteligencia artificial, las redes neuronales artificiales son modelos muy precisos en tareas como la clasificación y la regresión en el estudio de fenómenos naturales, pero se consideran “cajas negras” porque no permiten explicación directa de aquello que abordan. Este trabajo revisa la posibilidad de explicación científica a partir de estos modelos y concluye que se requieren de otros esfuerzos para entender su funcionamiento interno. Esto plantea retos para acceder a la explicación científica a través de su uso, pues la naturaleza de las redes neuronales artificiales dificulta a primera instancia la comprensión científica que puede extraerse de estas.

Palabras clave: redes neuronales artificiales; explicación científica; explicabilidad; interpretabilidad; transparencia; cajas negras.

* Este artículo se debe citar: Rodríguez-Sánchez, Alejandro E. “La posibilidad de explicación científica a partir de modelos basados en redes neuronales artificiales”. *Revista Colombiana de Filosofía de la Ciencia* 24.48 (2024): 161-194. <https://doi.org/10.18270/rcfc.4288>

ABSTRACT

In Artificial Intelligence, Artificial Neural Networks are very accurate models in tasks such as classification and regression in the study of natural phenomena, but they are considered “black boxes” because they do not allow direct explanation of what they address. This paper reviews the possibility of scientific explanation from these models and concludes that other efforts are required to understand their inner workings. This poses challenges to access scientific explanation through their use, since the nature of Artificial Neural Networks makes it difficult at first instance the scientific understanding that can be extracted from them.

Keywords: Artificial Neural Networks; Scientific Explanation; Explainability; Interpretability; Transparency; Black-boxes.

1. INTRODUCCIÓN

En las últimas dos décadas, los sectores de la ciencia y de la tecnología han visto desarrollo y avance respecto a un paradigma del conocimiento que, entendido como una amalgama de tecnologías y métodos, representa una alternativa con la cual predecir y estimar fenómenos sociales o naturales. Este paradigma se refiere al aprendizaje automático (AA), el cual subsume diversos grupos de modelos y algoritmos empleados generalmente para la predicción, la retrodicción y la estimación de un fenómeno de interés (Alzubaidi et ál. 2021; Dhall, Kaur & Juneja 2020; Emmert-Streib et ál. 2020). A la luz de los reportes científicos publicados y de la tecnología generada a partir de ellos, es justo decir que el AA ha sido continuamente impulsado tanto por la comunidad científica como por el sector tecnológico a nivel global para alcanzar sus respectivas metas: ganar más conocimiento y generar valor para un segmento del mercado (Saxe, Nelli & Summerfield 2021; Soniya, Paul & Singh 2015; Tomašev et ál. 2020; Wang, Zhao & Pourpanah 2020). El desarrollo del AA es, pues,

inherente a los beneficios que cosecha en diversos sectores, como el tecnológico y el industrial (Tang et ál. 2019).

Desde un nivel taxonómico alto, el AA pertenece al campo de la inteligencia artificial (IA), ya que los algoritmos, los métodos y las técnicas que en ello se emplean basan su funcionamiento en procesos informáticos, y estos, a su vez, se inspiran en procesos biológicos observados en la naturaleza, como lo es el proceso de aprendizaje y la adaptación a nueva información (LeCun, Bengio & Hinton 2015). Entre ellos se destacan los *modelos basados en redes neuronales artificiales*, de los cuales su funcionamiento guarda semejanza con el procesamiento interno que ocurre en los cerebros biológicos de los animales y de los seres humanos (Aggarwal 2018; Janiesch, Zschech & Heinrich 2021).

Las redes neuronales artificiales (RNA) se usan en espacios técnico-científicos —universidades, centros de investigación y empresas de tecnología— para múltiples aplicaciones y tareas, entre las que vale la pena destacar la predicción, el diagnóstico y el análisis de fenómenos naturales (Goodfellow, Bengio & Courville 2016). Más aún, se sabe que, por su adaptabilidad, la aplicación de las RNA tiene lugar en disciplinas tan disimiles unas de otras como la meteorología (e.g., Abhishek et ál. 2012), la astronáutica o la ciencia de los materiales (Faller & Schreck 1996; Ramprasad et ál. 2017; Silvestrini & Lavagna 2022; Song, Rondao & Aouf 2022). Sobre esto, se puede decir que dichas aplicaciones confirman el potencial y el valor que tienen estos modelos para la ciencia, la tecnología y la sociedad, de ahí el fuerte compromiso para entender su naturaleza. Como se verá, dicho valor repara precisamente en su elemento predictivo más que explicativo con respecto tanto a los objetos que se analizan desde sus elementos funcionales, como a sus propios algoritmos.

En los reportes relacionados con el quehacer de la profesión de la ingeniería (Chen & Liu 2022; Shehab et ál. 2022; Singh et ál. 2009), los modelos

basados en RNA actualmente demuestran potencia de predicción¹ muy superior respecto a otros modelos tradicionales. Esto les posiciona como herramientas de predicción y estimación por excelencia, puesto que su capacidad de abstracción de orden, forma y tendencia de los datos resulta conveniente para algunas aplicaciones y usos donde o bien los modelos existentes tienen un error alto, o no existen modelos que estimen adecuadamente la respuesta estudiada de un fenómeno. Por esto, en la praxis tecnológica y de la ingeniería, las RNA tienden a ser preferidas por sobre otros modelos, ya que son excelentes en lo que se considera *desiderata no-teórica* (Cichy & Kaiser 2019); esto es, se ejecutan rápido en un ordenador, son baratas de construir en contraposición con el tiempo de modelado matemático tradicional para estudios multifactoriales, y constituyen un uso de memoria informática eficiente una vez que están creadas y validadas con datos experimentales. Sin embargo, es importante destacar que el proceso de entrenamiento de las RNA también puede requerir un alto costo computacional y de tiempo, especialmente cuando se trabaja con grandes conjuntos de datos y arquitecturas complejas, como los modelos de lenguaje de gran tamaño².

Sin embargo, aun con las primacías que presentan las RNA en cuanto a objetivos de predicción, existe una cuestión relevante en su uso que reside en la explicación desde la perspectiva científica que se puede extraer a partir de ellas. Esto supone un problema relevante en las ciencias pues parece existir cautela en cuanto al valor explicativo que se puede lograr por medio de modelos basados en RNA para tareas de diseño, optimización y exploración de sistemas físicos (Zednik 2021). Este problema nace a raíz de la preocupación latente de la gran potencia de estimación y predicción que ofrecen las RNA para predecir fenómenos naturales sin que necesariamente se entienda, a primera instancia, su estructura interna o la relación que esta guardaría como parámetro explicativo de la regularidad de

¹ Potencia precisada en términos de los valores de medidas estadísticas que relacionan la distancia entre las predicciones de un modelo con respecto a datos experimentales.

² Del inglés *large language models*.

un fenómeno (Rudin 2019). Sobre ello, se podría argumentar que la inquietud es legítima, pues no son pocas las ocasiones en las que en el ambiente de las ciencias los científicos se decantan por el uso de un modelo con menores parámetros en los elementos que le conforman, en favor de la explicación que ofrecen las descripciones más sencillas, en lugar del uso de un modelo muy preciso pero complejo como lo suelen ser las RNA (principio metodológico conocido como *navaja de Occam*). El quid se ubica, entonces, en que para esbozar una explicación es requisito conocer el funcionamiento interno de un constructo lógico o matemático (Verreault-Julien 2019), lo cual en las RNA suele ser una tarea difícil y complicada si solo se aborda su construcción y aplicación, como está reportado en documentos técnicos y libros de texto dedicados a la enseñanza de estas³. Incluso en ámbitos no afines a las disciplinas relacionadas con la construcción de modelos de tal clase es común hacer alusión al término “caja negra” para referirse a la lógica de entrada-salida que permite condiciones y niveles de ingreso de uno o múltiples factores para emitir respuestas sin revelar su funcionamiento interno (Nathan 2021), como es el caso del *software* (Morin et ál. 2012).

Este trabajo presenta argumentación sobre los conceptos de *explicabilidad*, *interpretabilidad* y *transparencia* como ejes precursores de la explicación científica a partir de los modelos RNA del entorno del AA. Se abordarán los conceptos base de diferentes enfoques de explicación científica, siendo los modelos de cobertura legal de Hempel, el concepto de explicación de Van Fraassen (1980) y el modelo de causalidad de Pearl (2009) los mencionados para contextualizar la discusión de posibilidad de explicación científica a partir de modelos RNA. Con base en esto, el objetivo principal se centra en hacer una contribución a la educación en las ciencias y al sector de la tecnología respecto a lo que sucede en el

³ Estas serían dos etapas de rutina muy generales en la práctica y uso de dichos modelos, i. e., cosecha y ordenamiento de los datos de un fenómeno para el entrenamiento, la validación y la prueba del modelo en cuestión —o, más precisamente, su construcción—, mientras que la aplicación se refiere al despliegue o puesta en producción de estos modelos en alguna plataforma de tecnología. Desde luego, como se verá, cada una de ellas entraña una serie de actividades y subetapas que lleva a cabo el ingeniero o investigador que las conduce.

ámbito del entorno de modelado de problemas y fenómenos por medio de modelos basados en RNA, ya que actualmente es común que dichos constructos dominen el funcionamiento de plataformas tecnológicas que son usadas en la cotidianidad, las cuales día a día son aún más sofisticadas dado su continuo desarrollo, pero casualmente no son comprendidas a un nivel explicativo y por ello se generan problemáticas como la concepción de la caja negra. Así, lo que se pretende es que lo contenido aquí sirva para dar a conocer las bases de lo que significan dichos modelos, sus objetivos y, dado el actual avance tecnológico, la posible semejanza que ya pueden tener con otros constructos lógicos más parsimoniosos y transparentes, como los modelos analíticos de las ciencias, en cuanto a su capacidad para explicar un fenómeno.

Para poder abordar con suficiencia los conceptos que aquí se mencionan, se hace una revisión a las bases fundamentales de los modelos en las ciencias; se recurre a la definición de modelo científico, y se enfatiza en las concepciones nómicas de la explicación en su grado más general y en términos del *modelo de cobertura hempeliano de explicación científica*. Además, se señalan los esfuerzos que en la actualidad se hacen para ofrecer modelos explicativos de las RNA desde una perspectiva de la heurística computacional. Finalmente, se dedica una sección a la teoría esencial que describe los elementos matemáticos y la semántica de las RNA con el fin de presentar las partes fundamentales con las que estas están construidas.

2. MODELOS EN LA CIENCIA Y LA TECNOLOGÍA

La definición de los modelos científicos suele asociarse con réplicas de un objeto o con entidades de clase figurativa o formal de un fenómeno, y se construyen para entender parcelas de la realidad y del universo natural (Díaz 2005). En la actividad científica, estas vienen a ser representaciones abstractas e idealizadas

que consideran un rango de condiciones particulares para representar un fenómeno o un objeto (Frigg & Hartmann 2020). Así, en el ámbito de las ciencias, y como bien se apunta en el trabajo de Acevedo-Díaz et ál., “[...] el carácter ambiguo o polisémico de la idea de modelo, presente incluso dentro de la actividad científica en las diferentes disciplinas, [...] induce una comprensión incompleta, cuando no errónea del significado de modelo científico” (2017 157). Por ello es claro e importante destacar la definición del concepto de modelo científico para el caso de estudio de las RNA, en lo posible encerrando la significación de su relación con las actividades científicas.

La acepción de modelo científico como una representación de un objeto, comportamiento o fragmento de la realidad que se busca entender y explicar es acertada (Giere 2004)⁴. Sin embargo, se puede complementar agregando el término de interpretación, puesto que es coherente con el contexto del entendimiento de la naturaleza; es decir, según Bailer-Jones (2009), un modelo se refiere a una descripción interpretativa de un fenómeno que facilita el acceso a este. También, y en términos de sus objetivos, un modelo, según como lo describe el filósofo Carl G. Hempel (2005), sirve para la predicción, retrodicción y explicación de las regularidades del flujo de acontecimientos en la naturaleza. Bajo estos términos y acepciones, entonces, un modelo para el caso que ocupa en las ciencias es una representación codificada y abstraída de la realidad, de tal suerte que, una vez construidos, estos tienden a ser económicos en su aplicación por cuanto pueden ser explotados para diversos fines, como la predicción de un fenómeno natural o social, pero esencialmente apoyan para la consecución de una meta (por ejemplo, predecir el producto interno bruto de un país). No obstante, es importante señalar que, aunque los modelos permiten estimar estados de la regularidad de aquello que

⁴ Díaz en el 2005, por ejemplo, argumenta que la existencia de un modelo en las ciencias se da solo cuando hay una representación, en la que además se espera similitud con aquello que representa; esto es, “el modelo pretende constituirse en una analogía lo más cercana posible a una homología con las cosas, procesos o sistemas que interesan al estudioso, pero sin llegar a conseguirla” (Díaz 2005 13).

representan de la realidad, estos también cumplen la función de explicar cómo esa uniformidad da lugar a los hechos y las consecuencias que se desprenden de algunas condiciones observadas (Ladyman 2001).

Sobre la predicción y la retrodicción, estas corresponden con las actividades de estimar los estados futuros o pasados que presenta un fragmento de la realidad dentro de la regularidad que le supone. Así, por ejemplo, los modelos de la mecánica newtoniana nos permiten hacer inferencias de las posiciones de una roca después de ser arrojada desde lo alto de un acantilado a partir de su masa, sus condiciones iniciales y las condiciones de frontera del movimiento de esta (velocidades, aceleración, posición, región del movimiento, fuerzas que se le oponen). En el caso de las RNA, la retrodicción y predicción también se ciernen sobre los niveles de los factores o características de fenómeno.

El concepto de modelo en la tecnología se extiende a partir de las consideraciones previamente descritas desde la perspectiva en ciencia; es decir, los modelos en la tecnología también son constructos que formalizan la relación entre las condiciones y las respuestas de un fenómeno en particular, y, como tales, sirven para predecir y estimar alguna respuesta en concreto de este. La diferencia radica en que estos últimos se adhieren a fines que pueden ir más allá de la explicación de un fenómeno, como la predicción, el control o la intervención sobre este, y en el caso de la tecnología, dichos fines pueden incluir la generación de rentabilidad en una compañía. Además, aunque la taxonomía de modelos científicos y tecnológicos puede ser amplia, este artículo limita solo la mención y discusión de las RNA como “modelos computacionales” del tipo formal (Díaz 2005). Una descripción más profunda puede ser encontrada en el trabajo de Frigg y Hartmann (2020).

3. REDES NEURONALES ARTIFICIALES

Las redes neuronales artificiales son modelos computacionales del tipo formal (Díaz 2005). Estos se fundamentan en el trabajo de McCulloch y Pitts (1943) relacionado con el estudio del modelado lógico del comportamiento de una neurona biológica en sus funciones de recibir, procesar y emitir información. En sí, es justo decir que una sola de estas unidades no es capaz de generar inteligencia propia para modelar un fenómeno (Kandel et ál. 2013)⁵, de ahí que sea necesario contar con una red interconectada de estas para lograr el aprendizaje necesario y así representar datos como modelos científicos y tecnológicos (LeCun, Bengio & Hinton 2015). De hecho, las RNA deben su nombre a que son similares a los sistemas nerviosos de los seres vivos, en los que una red rica en neuronas conectadas entre sí da lugar a comportamientos y conductas complejas. En este contexto, la concepción de neurona artificial sería retomada con mayor énfasis y mejores resultados por Rosenblatt y otros investigadores durante las décadas de 1940 y 1950, como lo recopila Bishop (2015) en su revisión histórica. Rosenblatt llevó a cabo las primeras simulaciones de su “perceptrón” en la década de 1950, definido como un instrumento de procesamiento y aprendizaje de datos basado en entradas/salidas, y presentó un trabajo al respecto en 1958 (Rosenblatt 419-449). Posteriormente, en 1962, apareció su libro más importante sobre el tema, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Rosenblatt 1962). Consecuentemente, la forma moderna de las RNA ocurrió al incorporar el “algoritmo de retropropagación” de Rumelhart, Hinton y Williams (1986), y con dicho avance el perceptrón de Rosenblatt tendría la capacidad de

⁵ McCulloch y Pitts (1943), en su estudio, hacen una introducción y desarrollo de la lógica computacional que describe por primera vez las neuronas artificiales, pero no establecen términos que definan las redes neuronales artificiales.

representar información contenida en un conjunto de datos al incorporar elementos que relacionan entradas y salidas⁶.

De McCulloch, Pitts, Rosenblatt y Rumelhart se aduce que las RNA son modelos computacionales, ya que aprenden a representar aquello que se contiene en conjuntos de datos de información y sus propiedades son inherentes de la informática (los algoritmos que las constituyen se hacen posibles a través de procesamiento computacional). Aunque los trabajos de estos últimos autores mencionados refieren a los fundamentos y bases que dieron lugar a las RNA modernas, desde los años setenta se han visto diversas etapas de desarrollo, madurez y estacionalidad en cuanto a la ciencia que entraña el estudio de estas (LeCun, Bengio & Hinton 2015). En este artículo, la discusión y argumentación se centran en los modelos de RNA con “aprendizaje supervisado” a través del método de entrenamiento por retropropagación (O’Shea & Nash 2015; Schmidt 2019). En aras del estudio y la argumentación que se presenta en las próximas secciones, se prescinde de modelos RNA más sofisticados como los mencionados en los trabajos de Goodfellow, Bengio y Courville (2016) y Aggarwal (2018).

3.1. BREVE DESCRIPCIÓN MATEMÁTICA DE LAS RNA

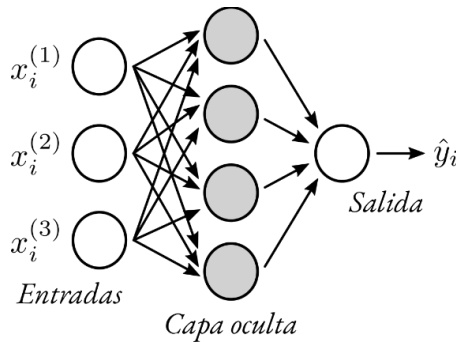
Las RNA no son solamente instrumentos y modelos que producen salidas a partir de entradas⁷. Como modelos formales, sus elementos se describen por medio de expresiones matemáticas y lógicas. Además, estas poseen unidades básicas llamadas “neuronas artificiales”, cada una de las cuales recibe y emite información si se alcanza un umbral definido por una función de activación. Precisamente, el

⁶ Dentro del contexto del AA y las RNA, no es para nada trivial que los textos de teoría especializados empleen la palabra inglesa *representation*, puesto que una RNA se asume propiamente como un modelo de aquella información que representa

⁷ Es particularmente importante resaltar que el entendimiento de estas en esos términos es el detonante de problemáticas del tipo caja negra en las RNA.

término red proviene de la conexión de varias neuronas, ya que por sí sola una no es capaz de capturar y representar las complejas relaciones y patrones presentes en los fenómenos que se busca modelar, de ahí que se requieran redes de estas (Kandel et ál. 2013). Así, el modelado mediante RNA implica la construcción de representaciones abstractas de los fenómenos estudiados a través del aprendizaje de patrones en los datos.

En una RNA de retropropagación, las neuronas artificiales están dispuestas por grupos llamados “capas” (del anglicismo *layers*), donde se alojan neuronas artificiales, y, a su vez, estas se conectan con otras capas a través de coeficientes que representan la intensidad de conexiones entre neuronas, también llamados “pesos”. La primera y la última capa se denominan “capa de entrada” y “capa de salida”, que, respectivamente, contienen las entradas y salidas de información de datos que pretenden procesar. A su vez, las capas intermedias se denominan “capas ocultas”, las cuales son las neuronas en las que sus pesos se ajustan para lograr un error de predicción aceptable en una RNA del tipo de retropropagación (figura 1).



*Nota: los superíndices indican la posición de una neurona y el subíndice i se refiere al i -ésimo ejemplar contenido en un conjunto de datos \mathcal{C} .

Figura 1. Diagrama de una red neuronal artificial de tres entradas, cuatro neuronas ocultas y una salida.

Fuente: elaboración propia.

La figura 1 muestra un esquema de una RNA con una sola capa oculta del tipo de retropropagación. Se describen las entradas por medio de un *vector de características*:

$$\mathbf{x}_i = \begin{bmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(m)} \end{bmatrix}, \mathbf{x}_i \in \mathbb{R}^m, \tag{1}$$

donde cada superíndice indica una característica del fenómeno que se pretende representar; el subíndice i representa el i -ésimo vector de características y \mathbb{R} se refiere al conjunto de los números reales (para el caso de la figura 1, $m = 3$). En este contexto, las características se entienden como las variables independientes sobre las cuales se da una correspondencia de la forma

$$f_{RNA}: \mathbb{R}^m \rightarrow \mathbb{R}^n, \tag{2}$$

donde f_{RNA} especifica un mapeo funcional desde \mathbb{R}^m hacia \mathbb{R}^n (para el caso concreto de la figura 1, $n = 1$). Esta correspondencia es precisamente la que un modelo de RNA exitoso representa, y se da a partir del aprendizaje de un conjunto de datos \mathcal{C} :

$$\mathcal{C} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_i^N, \mathbf{x}_i \in \mathbb{R}^m; \quad \mathbf{y}_i \in \mathbb{R}^n, \tag{3}$$

donde N es el tamaño del conjunto \mathcal{C} y \mathbf{y}_i es la respuesta o variable dependiente que se interpreta como la salida que se pretende modelar:

$$\mathbf{y}_i = \begin{bmatrix} y_i^{(1)} \\ y_i^{(2)} \\ \vdots \\ y_i^{(m)} \end{bmatrix}, \mathbf{y}_i \in \mathbb{R}^n. \quad (4)$$

Además, una RNA se construye cuando las conexiones internas entre cada una de sus neuronas, conocidas como “pesos”, se entrenan o calibran para efectivamente reducir el error existente entre sus salidas. Esto se hace a través de un proceso iterativo conocido como “descenso de gradiente”, el cual puede ser descrito por la siguiente expresión:

$$w_{l,k}^{(j)} \leftarrow w_{l,k}^{(j)} - \frac{\alpha}{N} \sum_{i=1}^N \left(\frac{\partial L}{\partial w_{l,k}^{(j)}} \right)_i, \quad (5)$$

donde $w_{l,k}^{(j)}$ se refiere al peso existente de la neurona k de la capa l con la neurona (j) de la capa previa ($l - 1$), esto es, define la relevancia de si una neurona recibe información para ser procesada. Además, α es una *tasa de aprendizaje* y la derivada parcial del lado derecho es el gradiente del error L para cada conexión:

$$L = \frac{1}{Nn} \sum_{\{(x_i, y_i)\}_{i=1}^N} \underbrace{\sum_{j=1}^n \left(y_i^{(j)} - \hat{y}_i^{(j)} \right)^2}_{C_i}, \quad (6)$$

donde C_i es un error individual para cada i -ésimo par en el conjunto de datos, y las predicciones o estimaciones de una RNA se alojan en el siguiente vector:

$$\hat{\mathbf{y}}_i = \begin{bmatrix} \hat{y}_i^{(1)} \\ \hat{y}_i^{(2)} \\ \vdots \\ \hat{y}_i^{(n)} \end{bmatrix}, \hat{\mathbf{y}}_i \in \mathbb{R}^n. \tag{7}$$

Las ecuaciones 1 a 7 describen los elementos de una RNA, por lo que, sucintamente, estas pueden ser circunscritas como un problema de optimización, es decir, el propósito de minimizar L a través de una heurística computacional, tal que

$$\min_{\mathbf{w}} L(\mathbf{w}); \quad L(\mathbf{w}) = \frac{1}{N} \sum_{i=1}^N C_i(\mathbf{w}), \tag{8}$$

donde w se refiere a la forma vectorizada de los pesos. De esta manera, en las RNA, el aprendizaje ocurre cuando los pesos⁸ se ajustan para reducir el error entre los datos y las predicciones a través de algoritmos iterativos. Uno de estos algoritmos es el descenso de gradiente, como el que se define por la ecuación 4.⁹

Como tal, los parámetros de una RNA no revelan nada del fenómeno que esta podría representar. Esto genera un problema desde la perspectiva científica pues, tal como se menciona en el estudio de Zednik del 2021, la preocupación se centra en que los parámetros de un modelo no ofrecen elementos para generar interpretación y explicación de los fenómenos que se buscan modelar a partir de un conjunto de datos. Es decir, contrario a un modelo matemático formal, aquí, por ejemplo, una matriz de pesos \mathbf{w} no contiene necesariamente valor epistémico

⁸ No son, sin embargo, los únicos parámetros que se modifican de un modelo; dependiendo de su arquitectura, una RNA podría también ver ajuste, por ejemplo, en otros parámetros b llamados “sesgos”.

⁹ Los algoritmos de actualización de los parámetros de una RNA se disponen en una taxonomía que les distingue como *heurísticos* y *metaheurísticos*; los primeros se basan en descripciones formales, similares a la ecuación 5, mientras que los segundos carecen de esto y recurren a lógica por analogía a algún proceso observado en la naturaleza, como los algoritmos genéticos (Yang 2010).

del fenómeno, aun si la predicción que se genera a partir de ello es muy precisa¹⁰. Esto adquiere mayor grado de opacidad cuando se dan arquitecturas de RNA del tipo profundo con más capas ocultas o neuronas de las que se muestran en la figura 1, es decir, modelos de RNA del *deep learning* (LeCun, Bengio & Hinton 2015) que requieren de millones de parámetros para su funcionamiento interno, pero que ninguno revela o dice algo sobre el fenómeno que aborda.

4. EL PROBLEMA DE LA CAJA NEGRA DE LOS MODELOS RNA

El término “caja negra”, en el caso semántico y práctico del modelado de fenómenos por RNA, implica que estas pueden cumplir con el objetivo de predecir y modelar fenómenos a partir de conjuntos de datos sin revelar lo que supone su

estructura interna para llevarlo a cabo (Zednik 2021). Esto es, las RNA no ofrecen a primera instancia interpretabilidad o transparencia en el proceso de predicciones o estimaciones, dificultando así la actividad de la explicación científica. Arriba se vio que los pesos y otros parámetros de una RNA nada dicen del fenómeno sobre el cual se construyen a partir de sus datos. Esto supone que el funcionamiento interno de estos modelos no permite establecer vínculo con la regularidad del fenómeno, provocando un problema de mayor trascendencia en las actividades dentro de las ciencias.

Según Rudin (2019), un modelo de caja negra bien puede ser una función que es demasiado complicada para ser entendida por un humano o un modelo reservado del cual no se tiene acceso a los mecanismos que le hacen funcionar. Las RNA caben dentro de esta definición a partir de lo que expresado

¹⁰ En este caso, me refiero a aquello que, en un modelo matemático tradicional, como lo es el caso de la masa m en la ecuación $\mathbf{F} = m\mathbf{a}$, está explícito y representa algo de aquello que tiene nombre, relevancia, e influencia del problema o del fenómeno que se estudia.

en la ecuación 2. Asociado a esto, tanto Rudin como otros autores señalan los riesgos de operar bajo estas condiciones, entre los que se destacan la posible toma de decisiones erróneas en instituciones públicas o en firmas especializadas conectadas a la economía (Barredo Arrieta et ál. 2020; Ghassemi et ál. 2021; Slack et ál. 2020). Esto se adhiere al principal problema que entraña la cuestión del término y concepto de caja negra, puesto que los modelos sirven, esencialmente, para tomar decisiones.

Muchos algoritmos de AA han sido nombrados modelos de caja negra debido a su inescrutable funcionamiento interno y la opacidad que existe en ellos (Barredo Arrieta et ál. 2020). En este sentido, dichos modelos no pueden servir para llevar a cabo explicaciones ya que cuando se busca una explicación lo que se pretende es entender por qué \hat{y}_i sucede a partir de ciertas condiciones establecidas en x_i ; es decir, y según Van Fraassen (1980), las explicaciones son respuestas a preguntas del tipo *¿por qué?*¹¹. Con base en esto, en un modelo que efectiva y positivamente es transparente, y que no se adhiere a la problemática de la caja negra (como lo es el caso de la *ley de los gases ideales*), sus elementos permiten positivamente establecer relaciones del tipo causa-efecto para responder a esta clase de interrogantes; por ejemplo, al aumentar la temperatura de un gas bloqueando la variable de volumen, se encuentra la razón de por qué se aumentaría la presión dentro de este¹². En esos términos, se podría plantear la idea de que en la actividad científica la legitimidad de un modelo de RNA tendría que ver con dos objetivos esenciales: (1) predecir y (2) explicar aquello que modela; en caso contrario, el problema de la caja negra se hace presente si no se interpreta y se da

¹¹ Un análisis más profundo de esta concepción la arroja el trabajo de Diéguez (1994), en el que se estudian los elementos que conforman una pregunta del tipo por qué: asunto, la clase de contraste y la relación de relevancia, siendo esta última lo que determina lo que contará como un factor explicativo a dicha pregunta.

¹² La forma más conocida de dicha ley está cifrada en la expresión matemática $PV = \nu RT$, donde P es la presión, V el volumen, ν es la cantidad de moles del gas, R la constante universal de los gases y T es la temperatura.

explicación de la relación entre las entradas, las salidas, y la relación causa-efecto de estas en un modelo.

Es importante enfatizar que, aunque a primera instancia las RNA adolecen de la parte explicativa, estas son excepcionales en la parte predictiva, más aún en la exploración y el mapeo de datos de los fenómenos que abordan (con ello es posible generar respuestas sintéticas de condiciones no testadas en las entradas de un modelo). Así, debe tenerse en cuenta el valor que aportan estos modelos, por lo que es esencial recurrir a métodos precursores que permitan a los investigadores llevar a cabo, primero, la tarea de transparentar el funcionamiento interno de un modelo, y, posteriormente, el trabajo de construcción de la explicación en términos científicos. Para abordar este problema es importante entender qué es la explicación en las ciencias, al menos desde un marco de entendimiento que haya sido o sea aceptado en la historia de la ciencia y la filosofía¹³.

5. MODELOS DE EXPLICACIÓN EN LAS CIENCIAS

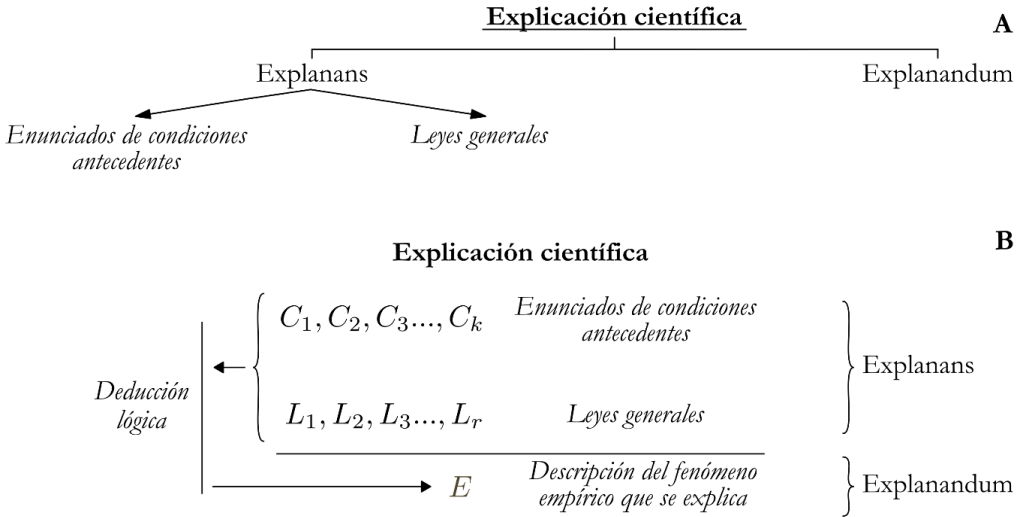
Como se ha planteado en la sección anterior, la explicación en las ciencias es responder a preguntas del tipo “por qué”. Aun cuando esa es, grosso modo, la forma sucinta y actualmente más aceptada del concepto, no es exclusiva de los modelos de explicación que en el pasado han servido de soporte a los científicos como marco de referencia para explicar las regularidades de un fenómeno en concreto. En esta sección se presentarán dos de las acepciones que han tenido mayor peso en la historia de la ciencia, lo cual permitirá abordar la discusión central de la existencia de explicación científica para el caso de las RNA.

¹³ Aquí debo hacer una aclaración: en la historia de la filosofía de la ciencia, la explicación científica se ha reconocido como una actividad fundamental en las ciencias. Existen múltiples abordajes y modelos mismos de esta clase de explicación. Sin embargo, los marcos que aquí se presentan establecen elementos que son válidos en muchos modelos matemáticos, lo que permite poner referencias para argumentar sobre la problemática de la caja negra en los modelos basados en RNA.

La figura 2(A) presenta un recuadro construido a partir del trabajo de Hempel, el cual refiere a su célebre modelo nomológico-deductivo (Hempel 2005). En dicho recuadro se pueden identificar dos partes esenciales del modelo de explicación en cuestión: (1) enunciados explanantes (*explanans*) y (2) enunciado *explanandum*. Estos dos elementos se describen a continuación:

1. El *explanandum* es la oración que describe el fenómeno a explicar (y no al fenómeno mismo).
2. El *explanans* son las clases de aquellas oraciones que se aducen para dilucidar el fenómeno.

Según Hempel (2005), las explicaciones son argumentos de dos tipos: (1) aquellos que indican las condiciones existentes y que se manifiestan antes de la explicación de un fenómeno, y (2) los que expresan o se subsumen en ciertas leyes generales del fenómeno. Además, según Hempel, la explicación en las ciencias debe cumplir dos requisitos esenciales para adquirir su estatus (2005): (i) tiene que ser relevante en términos explicativos (requisito de relevancia explicativa), y (ii) debe ser empíricamente contrastable (esto es, debe cumplir con el requisito de contrastabilidad).



(A) Descripción de sus elementos. (B) Deducción lógica de la explicación
Figura 2: Modelo de cobertura de la explicación científica de Hempel
Fuente: Hempel 2005.

Muchas explicaciones de fenómenos físicos caben en el modelo nomológico-deductivo, ya que es por medio de modelos matemáticos construidos a través de leyes generales físicas que es posible explicar el fenómeno que estos abordan. Esto se observa en la figura 2(B), en la que se señala que los *explanans* son las condiciones de las características de un fenómeno, las cuales se cifran y se relacionan entre sí a través de la instancia de un modelo, que es construido a partir de una o varias leyes generales. Sobre esto, tómese, por ejemplo, el caso particular de un cilindro de gas expuesto a altas temperaturas; aquí la ley y las condiciones, como *explanans*, permitirían explicar el comportamiento del incremento en su presión interna (*explanandum*).

Por otra parte, aunque pareciese tener una cobertura amplia, el mismo modelo hempeliano de explicación encuentra límites, como lo describe su creador, pues en su trabajo (Hempel 2005) se menciona que muchos fenómenos no solo se subsumen en una primera regularidad definida sobre alguna ley física, sino que también lo hacen en regularidades propias de la estocasticidad. En este sentido, el modelo inductivo-estadístico de Salmon (1970) surge como una alternativa al modelo de Hempel, al considerar que las explicaciones científicas pueden basarse en leyes estadísticas y no solo en leyes deterministas. Según Salmon, una explicación estadística es adecuada si el fenómeno a explicar se puede subsumir bajo una ley estadística de alto nivel. Es importante mencionar esto ya que toda RNA es construida a partir de datos que en muchas ocasiones siguen una tendencia general, al mismo tiempo que presentan una marcada variabilidad en su contenido, lo que las hace más compatibles con un enfoque estadístico de la explicación científica.

Existe otro abordaje más reciente de la explicación en las ciencias que viene soportado por los *modelos causales de explicación científica*, los cuales fueron desarrollados a finales del siglo XX e inicios del siglo XXI en pleno auge de los modelos de la inferencia estadística y la computación. Estos modelos son marcos que se usan para representar las relaciones entre las características de un sistema y las causas y los efectos de los cambios que estas sufren cuando hay un cambio en ellas. Como tal, la idea principal de dichos modelos es comprender la estructura de causas en un sistema para hacer predicciones, entender los fenómenos que se estudian y así poder dar explicarlos. El marco de modelos de causalidad más relevante es el “modelo de causalidad estructural” propuesto por Pearl y Mackenzie (2018). Este proporciona un conjunto de reglas y algoritmos para inferir la estructura de causas en un sistema. Según los autores, la explicación surge cuando los seres humanos hacen razonamiento contrafactual, el cual genera escenarios del tipo “¿qué podría ocurrir o habría ocurrido si...”. En ellos se puede entender las posibles consecuencias de haber seguido o considerado diferentes

niveles en las características de un sistema, aun si no se cuenta con datos de este (de ahí que sean contrafácticos)¹⁴.

Aunque el modelo nomológico-deductivo y los modelos de causalidad no son a primera instancia compatibles, sí es importante señalar que ambos, incluso en épocas distintas, han gozado de aceptación por su acometido para el desarrollo de la explicación de un fenómeno. Así, aunque de valor epistémico distinto, estos enfoques ponen en relieve dos situaciones clave para el caso de la explicación: (1) para ejercer una explicación del tipo científico se requieren leyes que cubran el fenómeno; (2) se necesitan llevar a cabo razonamientos del tipo contrafactual para establecer las hipótesis de la causa de un fenómeno. Ambas situaciones parecieran no ser posibles cuando se tiene un modelo RNA, puesto que el mero hecho de que permita predecir con excelente precisión una manifestación natural no significa que a partir de este se puedan plantear contrafactuals, ni que sus elementos básicos pertenecen a un conjunto dado de datos que modelan para construir un *explanandum*; esto es, las neuronas artificiales y algoritmos de entrenamiento no pertenecen a una ley que lo abarca todo (tómese por ejemplo el caso de si se modelara una instancia cubierta por la *ley de los gases ideales*). En esencia, esto es lo que no establece un vínculo posible para que de una RNA poder llevar a cabo el proceso de la explicación en un primer intento. De ahí que en tiempos recientes se han depositado esfuerzos desde diferentes marcos y frentes para establecer métodos y herramientas como precursores de la interpretación de un modelo opaco como lo son las RNA.

¹⁴ En el caso del cilindro de gas, es interesante preguntarse qué habría pasado con la temperatura si esta no se hubiera incrementado.

6. EXPLICABILIDAD, INTERPRETABILIDAD Y TRANSPARENCIA EN EL AA

Los filósofos llevan siglos discutiendo y ampliando el concepto de explicación en las ciencias (Woodward & Ross 2021), pero en el campo de los datos y el AA recién se comienzan a desarrollar términos sui generis que dan fundamento para tratar de explicar modelos como las RNA (Linardatos, Papastefanopoulos & Kotsiantis 2021). Como se ha visto, un modelo de este tipo es ciertamente opaco, y para poder aspirar a llevar a cabo ejercicios contrafactuales o de desarrollo de *explanandum* de un fenómeno, es necesario primero entender qué mecanismos de su interior hacen que sea funcional.

Existen tres conceptos clave dentro del AA que aluden a explicación, pero que no lo son si se limita este término al caso de explicar el fenómeno que aborda un modelo:

- *Interpretabilidad*. Los científicos de datos e ingenieros del AA le llaman “interpretabilidad” a la capacidad de explicar o presentar en términos comprensibles la información interna de un modelo a un humano (Doshi-Velez & Kim 2017). Según Vilone y Longo (2021), la interpretabilidad es el mapeo de un concepto abstracto (por ejemplo, una clase estimada) en un dominio al que el humano puede dar sentido. Un modelo interpretable podría ser un modelo de árbol de decisiones simple, del cual puede entenderse su funcionamiento tan solo con observar su estructura. De esta manera, se identifica que la interpretabilidad corresponde a (1) un examen del propio modelo o (2), alternativamente, a un examen de la respuesta que presenta un modelo a estímulos planificados que se llevan a cabo al considerar diferentes niveles en sus entradas (incluso de naturaleza estocástica).

- *Explicabilidad.* Se refiere a una colección de artefactos visuales o interactivos, como los gráficos, que proporcionan al usuario de un modelo de IA una descripción suficiente de su comportamiento; en los términos expuestos por Saleem et ál. (2022), la explicabilidad se trata de explicar el proceso interno de un modelo para responder a la pregunta “¿Cómo toma determinadas decisiones el modelo de caja negra de la IA?”.
- *Transparencia.* Según Roscher et ál. (2020), se dice que un enfoque de AA se hace transparente si los procesos que extraen los parámetros del modelo a partir de los datos de entrenamiento y generan etiquetas a partir de los datos de prueba pueden ser descritos y motivados por el diseñador del enfoque. Como tal, la transparencia se refiere a la arquitectura general de una RNA (o, en general, de un modelo de AA), sus componentes individuales del modelo, el algoritmo de entrenamiento que se usó para calibrarla y cómo se obtienen las soluciones mediante su uso (Sokol et ál. 2022).

Es importante exponer que los tres conceptos de explicabilidad tienen una connotación y trasfondo eminentemente informáticos. Esto es, a excepción del uso práctico que ofrece la interpretabilidad, estos conceptos no tienen un significado más allá de la explotación y consecuente aplicación de un modelo RNA que se da sobre una plataforma computacional. Más aún, en la ausencia de ello pierden sentido, pues estos se constituyen solo cuando existe y se da un modelo que ocupa espacio tangible en memoria electrónica. Por ello es importante aclarar que estos conceptos más bien se refieren a metamodelos¹⁵ de una RNA —

¹⁵ Según Van Gigch (1991), un metamodelo es un modelo que incorpora los elementos y las condiciones de otro modelo cuando este último posee características que lo hacen de difícil acceso; es decir, un metamodelo es un modelo sustituto de un modelo. Según este autor, los metamodelos se utilizan cuando los resultados de un modelo son demasiado complejos e intrincados, demasiado costosos de producir o no es factible trabajar con ellos. En el caso de los metamodelos de explicación, estos son modelos sustitutos que se usan para interpretar un modelo intrincado, como una RNA.

o de alguna otra clase de modelo informático del AA —¹⁶. Estos tres conceptos, además, se alejan de toda formalidad como la que ofrece Hempel en su modelo nomológico-deductivo. Esto no es un rasgo mismo de la disciplina en la que ven su desarrollo, más bien es una consecuencia del avance en construcción de nuevas tecnologías; es decir, son modelos informáticos que explican los modelos de IA, como las RNA. De esta manera, un metamodelo de explicabilidad es un modelo que explica el interior de un modelo opaco, no un metamodelo que explica la regularidad de un fenómeno.

Los términos de interpretabilidad, explicabilidad y transparencia tampoco se adhieren a los modelos de causalidad puesto que al ejecutarlos para transparentar y entender solo el funcionamiento de un modelo en concreto estos métodos no son para explicar aquello que abarca el modelo; en términos de Pearl y Mackenzie (2018 6), los “datos son mudos”¹⁷ y, dado que un modelo RNA se construye a partir de estos, no se podría esperar que los métodos relativos a la explicabilidad sean sinónimo de la explicación científica de la regularidad que describen tales datos. Lo que se entiende como explicabilidad, en el contexto informático, consiste, entonces, en un modelo mismo de los parámetros internos de una RNA a suerte de traducción de lo que ocurre en su interior. Así, las acepciones de interpretabilidad, explicación y transparencia en la ciencia de datos están ligadas al análisis de la relación entradas/salidas de un modelo, generalmente mediante técnicas como las librerías LIME y los modelos agnósticos (Ribeiro, Singh & Guestrin 2016); los modelos de SHAP (Lundberg & Lee 2017); las gráficas de efecto local acumulado ALE (Apley & Zhu 2020); la importancia de características (Guyon & Elisseeff 2003), o técnicas de graficado por dependencia

¹⁶ Al constituirse como herramientas informáticas, tanto los metamodelos de explicabilidad como los modelos de interpretabilidad tienen lugar toda vez que primero se ajusta y entrena un modelo de AA, en este caso una RNA, a través de sus predicciones. En el caso de la transparencia, esta se constituye si y solo si existe un modelo ya construido y validado.

¹⁷ Traducido por el autor directamente de la frase “data is profoundly dumb”, expuesta en el trabajo *The Book of Why* (Pearl & Mackenzie 2018 6).

parcial (Greenwell 2017). Estas herramientas son técnicas y modelos de modelos (metamodelos), de ahí su relevancia en cuestiones prácticas.

7. POSIBILIDAD DE EXPLICACIÓN A PARTIR DE UN MODELO RNA

Como se ha visto en secciones precedentes, una RNA se constituye a partir de conjuntos de datos, por lo que, con precisión, estos modelos también son modelos de datos¹⁸, esto es, representan un conjunto de datos que se obtiene a partir de la observación y la experimentación. Sin embargo, las RNA no son interpretables a primera instancia y se requieren metamodelos de explicabilidad que desentrañen su funcionamiento interno. Además, estos modelos tienen un valor práctico notable, de ahí su auge, pues la utilidad que proveen es para predecir y estimar datos con una elevada exactitud. Sin embargo, a la luz de lo planteado en las secciones anteriores, no se puede explicar un fenómeno a través de las RNA si solo se las tiene en cuenta, como sí ocurre en el caso de otros modelos en las ciencias (por ejemplo, en los modelos matemáticos).

El término de explicación en el contexto de explicabilidad a través de metamodelos, entonces, no hace una referencia directamente a explicar el fenómeno que una RNA ajusta para las tareas y los propósitos ya mencionados. El término de explicación, en el contexto de las ciencias, aborda algo más amplio y trascendente que el solo hecho de desentrañar el interior de un modelo, esto es, entendimiento científico, y es lo que plantean Krenn et ál. (2022). Por esto, también es justo por extensión decir que los metamodelos de explicación no hacen

¹⁸ Aquí es importante destacar que me refiero a la definición tal como se menciona en el trabajo de Frigg y Hartmann (2020), y que refiere al concepto de Suppes (1962), según el cual un modelo de datos es una versión corregida, rectificadora, reglamentada y en la mayoría de los casos idealizada de los datos que se obtienen a partir de la observación inmediata, es decir de los llamados datos brutos.

ni generan por sí mismos contrafactuales para poder ofrecer la explicación del fenómeno que aborda un modelo RNA. Como tal, la explicabilidad y la interpretabilidad se pueden entender como “métodos precursores” que sirven de soporte para una posterior explicación científica si se habla de su uso en redes neuronales. Sin embargo, tanto las RNA como estos metamodelos pueden no ser suficientes por sí solos para alcanzar un entendimiento científico; la explicabilidad solo es la capacidad de describir la relación entre la entrada y la salida de un modelo de forma comprensible para un ser humano, y la interpretabilidad se refiere a la capacidad de comprender su funcionamiento interno. Por esto, el mero hecho de poder explicar o interpretar el comportamiento interno de una RNA, incluso de manera no formal (ausencia de descripción matemática), no garantiza una explicación científica, ya que los propios modelos no se basan en los principios y las relaciones funcionales que tratan de aprender o representar, y estos pueden no coincidir con conocimientos científicos previos. Un reducto excesivo es el siguiente: incluso esos modelos de explicabilidad requieren ser interpretados por alguien. Como siempre lo ha sido, es tarea de los científicos interpretar los constructos que se erigen a partir de observaciones y luego explicar a partir de estos. Nada de eso ha cambiado en la actualidad.

Como se plantea en un trabajo reciente de Krenn et ál. (2022), y siguiendo las ideas planteadas por De Regt y Dieks (2005), para que un modelo de IA pudiese generar entendimiento científico y, por extensión, una explicación de tal categoría, este tendría que cumplir dos condiciones específicas:

1. Reconocer cualidades y características de una teoría sin realizar cálculos exactos, así como utilizarlas en un nuevo contexto.
2. Adquirir conocimientos científicos y transferirlos a un humano.

Según dichos autores, esto quiere decir que solo se está en condiciones de comprender un fenómeno cuando existe tal teoría inteligible de este que los cien-

tíficos reconocen cualitativamente consecuencias y características de esta sin necesariamente profundizar en cálculos exactos (como ejemplo téngase en cuenta el caso de la *ley de los gases ideales* ya descrita arriba, en la que no hubo necesidad de generar nuevos datos a partir de experimentación para llegar a la conclusión mencionada). Con base en lo anterior, una explicación científica está ligada al entendimiento científico, y para que esto ocurra debe existir una teoría (o los cimientos de esta)¹⁹, cosa que puede estar ausente en un modelo de datos, como ya se ha descrito en las secciones precedentes. Cuando eso sucede, ningún método automático computacional podría brindar una plena explicación científica sin un usuario que lo explote y lo interprete, ni sin una teoría. Es decir, es tarea del usuario de estos métodos generar una explicación de la regularidad que en principio pretende modelar con un modelo basado en redes neuronales artificiales, y que es precisamente el objetivo que se persigue tras recurrir a los métodos precursores de la explicabilidad en la mayoría de los modelos del AA.

8. REFLEXIONES FINALES

La explicación en la ciencia de datos y el AA por medio de técnicas y modelos computacionales está en pleno desarrollo, lo cual inexorablemente incluye a las RNA. Aunque en la actualidad hay reportes que dan evidencia de la posibilidad de simbolización matemática para interpretabilidad de modelos RNA mediante otras técnicas de metamodelos (ejemplos muy concretos de esto pueden ser consultados en los trabajos de Abdusalamov et ál. 2023, Alaa & Van der Schaar 2019, Vedantam et ál. 2019), lo cierto es que no hay una regla, una técnica o un método general aplicable a todos los casos (estos avances son atomizados, por

¹⁹ Aquí me refiero a modelos incipientes que no abarcan múltiples condiciones, también a la taxonomía y clases que definen a un problema, al igual que reportes de aquello sobre lo que se espera se logre la madurez para construir una teoría.

área específica y los métodos que reportan no son generalizables a otros casos todavía). Esta situación plantea la cuestión de si es posible o no alcanzar una regla o método general en ese campo, y si este desafío es exclusivo de la ciencia de datos y el AA o si se presenta también en otros ámbitos científicos. Si bien dicha pregunta merece un abordaje más profundo que excede el alcance del presente trabajo, es importante tener en cuenta su relevancia al considerar los desafíos y las limitaciones actuales de la explicación en modelos computacionales. A pesar de considerarse formales, no se debe perder la noción de que un modelo computacional de datos conlleva opacidad y requiere de un esfuerzo extra de sus usuarios para poder establecer ese vínculo que solo la explicación científica permite en la comprensión de un fenómeno.

Por otra parte, para efectivamente considerar el estatus de explicación científica a partir de modelos RNA, se requiere de un conjunto de técnicas y métodos precursores, una teoría del fenómeno que se aborda (o los cimientos de ella) y un entendimiento científico de este. Dichos elementos son fundamentales para alcanzar tal condición en las ciencias, como se ha discutido en los estudios citados a lo largo de las secciones precedentes.

Las redes neuronales artificiales y los modelos más complejos que se construyen con estas, aunque presentan problemas para una explicación científica plena, sí son prácticos, pues, como se plantea en Krenn et ál. (2022 763), se aducen como “microscopios computacionales”, los cuales “permiten investigar objetos o procesos que no pueden visualizarse o sondearse de ninguna otra manera, por ejemplo, procesos biológicos, químicos o físicos que ocurren a escalas de longitud y tiempo no accesibles en los experimentos”. Este valor se reconoce en la actualidad, y ha llevado al sector tecnológico a adoptarlas para resolver problemas de clase multifactorial, donde generalmente es difícil encontrar un modelo tradicional para resolver problemas (Ahmed, Wahed & Thompson 2023), o en el caso de las ciencias, para explorar soluciones de la respuesta de un sistema natural y, por subsiguiente, tratar de explicarlo como actividad humana.

REFERENCIAS

- Abdusalamov, Rasul et ál. “Automatic Generation of Interpretable Hyperelastic Material Models by Symbolic Regression”. *International Journal for Numerical Methods in Engineering* (2023): 1-12.
- Abhishek, Kumar et ál. “Weather Forecasting Model Using Artificial Neural Network”. *Procedia Technology* 4 (2012): 311-318. <<https://doi.org/10.1016/j.protcy.2012.05.047>>
- Acevedo-Díaz, José Antonio et ál. “Modelos científicos: significado y papel en la práctica científica”. *Revista Científica* 30.3 (2017): 155-166. <<https://doi.org/10.14483/23448350.12288>>
- Aggarwal, Charu C. *Neural Networks and Deep Learning*. Cham: Springer, 2018. <<https://doi.org/10.1007/978-3-319-94463-0>>
- Ahmed, Nur, Wahed, Muntasir y Thompson, Neil C. “The growing influence of industry in AI research”. *Science* 379 (2023): 884-886. <<https://doi.org/10.1126/science.ade2420>>
- Alaa, Ahmed M. y Van der Schaar, Mihaela. “Demystifying black-box models with symbolic metamodels”. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2019. 1-11.
- Alzubaidi, Laith et ál. “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions”. *Journal of Big Data* 8.1 (2021): 53.
- Apley, Daniel W. y Zhu, Jingyu. “Visualizing the effects of predictor variables in black box supervised learning models”. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 1 (2020): 1059-1086.
- Bailer-Jones, Daniela M. *Scientific Models in Philosophy of Science*. Pittsburgh: University of Pittsburgh Press, 2009.
- Barredo Arrieta, Alejandro et ál. “Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI”. *Information Fusion* 58 (2020): 82-115. <<https://doi.org/10.1016/j.inffus.2019.12.012>>

- Bishop, J. Mark. "History and philosophy of neural networks". *Computational Intelligence, Volume I*. Hisao Ishibuchi (ed.). París: EOLSS Publications, 2015.
- Chen, Jie y Liu, Yongming. "Fatigue modeling using neural networks: A comprehensive review". *Fatigue & Fracture of Engineering Materials & Structures* 45.4 (2022): 945-979.
- Cichy, Radoslaw M. y Kaiser, Daniel. "Deep neural networks as scientific models". *Trends in Cognitive Sciences* 23.4 (2019): 305-317. <<https://doi.org/10.1016/j.tics.2019.01.009>>
- De Regt, Henk y Dieks, Dennis W. "A contextual approach to scientific understanding". *Synthese* 144.1 (2005): 137-170. <<https://doi.org/10.1007/s11229-005-5000-4>>
- Dhall, Devanshi, Kaur, Ravinder y Juneja, Mamta. "Machine learning: A review of the algorithms and its applications". *Intelligent Computing and Applications*, Pradeep Kumar Singh et ál. (eds.). Cham: Springer International Publishing, 2020. 47-63.
- Díaz, José Luis. "Modelo científico: conceptos y usos". *El modelo en la ciencia y la cultura*. Alfredo López-Austin (ed.). Ciudad de México: Siglo XXI Editores/Universidad Nacional Autónoma de México, 2005. 11-28.
- Diéguez, Antonio J. "Explicando la explicación". *Daimon Revista Internacional de Filosofía* 8 (1994): 83-108. <<https://doi.org/10.58680/ej19947514>>
- Doshi-Velez, Finale y Kim, Been. "Towards a rigorous science of interpretable machine learning". *arXiv: Machine Learning* (2017): 1-13.
- Emmert-Streib, Frank et ál. "An introductory review of deep learning for prediction models with big data". *Frontiers in Artificial Intelligence* 3 (2020): 4.
- Faller, William E. y Schreck, Scott J. "Neural networks: Applications and opportunities in aeronautics". *Progress in Aerospace Sciences* 32.5 (1996): 433-456. <[https://doi.org/10.1016/0376-0421\(95\)00011-9](https://doi.org/10.1016/0376-0421(95)00011-9)>
- Frigg, Roman y Hartmann, Stephan. "Models in Science". *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta. (ed.). Spring/Metaphysics Research Lab, Stanford University, 2020. *Online*. <<https://plato.stanford.edu/archives/spr2020/entries/models-science/>>

- Ghassemi, Marzyeh et ál. “The false hope of current approaches to explainable artificial intelligence in health care”. *The Lancet Digital Health* 3.11 (2021): 745-750. <[https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9)>
- Giere, Ronald N. “How models are used to represent reality”. *Philosophy of Science* 71.5 (2004): 742-752. <<https://doi.org/10.1086/425063>>
- Goodfellow, Ian, Bengio, Yoshua y Courville, Aaron. *Deep Learning*. Cambridge: The MIT Press, 2016.
- Greenwell, Brandon M. “pdp: An R package for constructing partial dependence plots”. *The R Journal* 9 (2017): 421-436.
- Guyon, Isabelle y Elisseeff, André. “An introduction to variable and feature selection”. *Journal of Machine Learning Research* 3 (2003): 1157-1182.
- Hempel, Carl G. *La explicación científica: estudios sobre la filosofía de la ciencia*. Barcelona: Paidós Surcos, 2005.
- Janiesch, Christian, Zschech, Patrick y Heinrich, Kai. “Machine learning and deep learning”. *Electronic Markets* 31.3 (2021): 685-695. <<https://doi.org/10.1007/s12525-021-00475-2>>
- Kandel, Eric R. et ál. *Principles of Neural Science*. 5ª ed. Nueva York: McGraw-Hill Education, 2013.
- Krenn, Mario et ál. “On scientific understanding with artificial intelligence”. *Nature Reviews Physics* 4.12 (2022): 761-769. <<https://doi.org/10.1038/s42254-022-00518-3>>
- Ladyman, James. *Understanding Philosophy of Science*. Londres: Routledge, 2001.
- LeCun, Yann, Bengio, Yoshua y Hinton, Geoffrey. “Deep learning”. *Nature* 521.7553 (2015): 436-444. <<https://doi.org/10.1038/nature14539>>
- Linardatos, Pantelis, Papastefanopoulos, Vasilis y Kotsiantis, Sotiris. “Explainable AI: A review of machine learning interpretability methods”. *Entropy* 23.1 (2021): 18. <<https://doi.org/10.3390/e23010018>>
- Lundberg, Scott M y Lee, Su-In. “A unified approach to interpreting model predictions”. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc., 2017. 4768-4777.
- McCulloch, Warren S y Pitts, Walter. “A logical calculus of the ideas immanent in nervous activity”. *The Bulletin of Mathematical Biophysics* 5.4 (1943): 115-133. <<https://doi.org/10.1007/BF02478259>>

- Morin, Alexander et ál. “Shining light into black boxes”. *Science* 336.6078 (2012): 159-160. <<https://doi.org/10.1126/science.1218263>>
- Nathan, Marco J. *Black Boxes: How Science Turns Ignorance into Knowledge*. New York: Oxford University Press, 2021.
- O’Shea, Keiron y Nash, Ryan. “An introduction to convolutional neural networks”. *arXiv abs/1511.08458* 1 (2015): 1-12.
- Pearl, Judea. *Causality: Models, Reasoning and Inference*. 2ª ed. Nueva York: Cambridge University Press, 2009.
- Pearl, Judea y Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Nueva York: Basic Books, Inc., 2018.
- Ramprasad, Rampi et ál. “Machine learning in materials informatics: recent applications and prospects”. *npj Computational Materials* 3.1 (2017): 54.
- Ribeiro, Marco Tulio, Singh, Sameer y Guestrin, Carlos. “Model-agnostic interpretability of machine learning”. *arXiv 1602.04938* (2016): 1-5.
- Roscher, Ribana et ál. “Explainable machine learning for scientific insights and discoveries”. *IEEE Access* 8 (2020): 42200-42216. <<https://doi.org/10.1109/ACCESS.2020.2976199>>
- Rosenblatt, Frank. “Two theorems of statistical separability in the perceptron”. *The Mechanisation of Thought Processes: Proceedings of a Symposium Held at the National Physical Laboratory*. Vol. 1. Londres: HMSO, 1958. 419-449.
- _____. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Ann Arbor: Spartan Books, 1962.
- Rudin, Cynthia. “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”. *Nature Machine Intelligence* 1.5 (2019): 206-215.
- Rumelhart, David E., Hinton, Geoffrey E. y Williams, Ronald J. “Learning representations by back-propagating errors”. *Nature* 323.6088 (1986): 533-536. <<https://doi.org/10.1038/323533a0>>
- Saleem, Rabia et ál. “Explaining deep neural networks: A survey on the global interpretation methods”. *Neurocomputing* 513 (2022): 165-180. <<https://doi.org/10.1016/j.neucom.2022.09.129>>
- Salmon, Wesley C. “Statistical explanation”. *The Nature and Function of Scientific Theories*. Robert Colodny (ed.). Pittsburgh: University of Pittsburgh Press, 1970. 173-231.

- Saxe, Andrew., Nelli, Stephanie., y Summerfield, Christopher. "If deep learning is the answer, what is the question?". *Nature Reviews Neuroscience* 22.1 (2021): 55-67. <<https://doi.org/10.1038/s41583-020-00395-8>>
- Schmidt, Robin M. "Recurrent Neural Networks (RNNs): A gentle introduction and overview". *arXiv abs/1912.05911* (2019): 1-16.
- Shehab, Mohammad et ál. "Artificial neural networks for engineering applications: a review". *Artificial Neural Networks for Renewable Energy Systems and Real-World Applications*. Ammar H. Elsheikh y Mohamed Elasyed Abd Elaziz (eds.). Academic Press, 2022. 189-206.
- Silvestrini, Stefano y Lavagna, Michèle. "Deep learning and artificial neural networks for spacecraft dynamics, navigation and control". *Drones* 6.10 (2022). <<https://doi.org/10.3390/drones6100270>>
- Singh, Yogesh et ál. "Application of neural networks in software engineering: A review". *Information Systems, Technology and Management*. Sushil K. Prasad et ál. (eds.). Berlín/Heidelberg: Springer Berlin Heidelberg, 2009. 128-137.
- Slack, Dylan et ál. "Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods". *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. Nueva York, NY: Association for Computing Machinery, 2020. 180-186.
- Sokol, Kacper et ál. "What and how of machine learning transparency: Building bespoke explainability tools with interoperable algorithmic components". *Journal of Open Source Education* 5.58 (2022): 175. <<https://doi.org/10.21105/jose.00175>>
- Song, Jianing, Rondao, Duarte y Aouf, Nabil. "Deep learning-based spacecraft relative navigation methods: A survey". *Acta Astronautica* 191 (2022): 22-40. <<https://doi.org/10.1016/j.actaastro.2021.10.025>>
- Soniya, Paul, Sandeep y Singh, Lotika. "A review on advances in deep learning". *2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*. IEEE (2015): 1-6. <<https://doi.org/10.1109/WCI.2015.7495514>>
- Suppes, Patrick. "Models of data". *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*. Ernest Nagel et ál. (eds.). Stanford University Press, 1962. 252-261.

- Tang, Binhua et ál. “Recent advances of deep learning in bioinformatics and computational biology”. *Frontiers in Genetics* 10 (2019): 214.
- Tomašev, Nenad et ál. “AI for social good: Unlocking the opportunity for positive impact”. *Nature Communications* 11.1 (2020): 2468.
<<https://doi.org/10.1038/s41467-020-15592-3>>
- Van Gigch, John P. *System Design Modeling and Metamodeling*. Nueva York: Springer Science, 1991.
- Van Fraassen, Bas C. *The Scientific Image*. Oxford: Clarendon Press, 1980.
<<https://doi.org/10.1093/0198244274.001.0001>>
- Vedantam, Ramakrishna et ál. “Probabilistic neural symbolic models for interpretable visual question answering”. *Proceedings of the 36th International Conference on Machine Learning*. Vol. 97. Kamalika Chaudhuri y Ruslan Salakhutdinov (ed.). Red Hook, NY: Curran Associates Inc., 2019. 6428-6437.
- Verreault-Julien, Philippe. “How could models possibly provide how-possibly explanations?”. *Studies in History and Philosophy of Science Part A* 73 (2019): 22-33. <<https://doi.org/10.1016/j.shpsa.2018.06.008>>
- Vilone, Giulia y Longo, Luca. “Notions of explainability and evaluation approaches for explainable artificial intelligence”. *Information Fusion* 76 (2021): 89-106. <<https://doi.org/10.1016/j.inffus.2021.05.009>>
- Wang, Xizhao, Zhao, Yanxia y Pourpanah, Farhad. “Recent advances in deep learning”. *International Journal of Machine Learning and Cybernetics* 11.4 (2020): 747-750. <<https://doi.org/10.1007/s13042-020-01096-5>>
- Woodward, James y Ross, Lauren. “Scientific explanation”. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta. (ed.). Summer/Metaphysics Research Lab, Stanford University, 2021. *Online*.
<<https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation/>>
- Yang, Xin-She. *Engineering Optimization: An Introduction with Metaheuristic Applications*. Hoboken: Wiley, 2010.
- Zednik, Carlos. “Solving the black box problem: A normative framework for explainable artificial intelligence”. *Philosophy & Technology* 34.2 (2021): 265-288.

ENTIDADES FÍSICAS, IDENTIDAD TEMPORAL Y BASURA ESPACIOTEMPORAL. UNA RECONSTRUCCIÓN CRÍTICA DE LA ONTOLOGÍA FÍSICA DE DOWE*

PHYSICAL ENTITIES AND SPATIOTEMPORAL JUNK. A CRITICAL RECONSTRUCTION OF P. DOWE'S PHYSICAL ONTOLOGY

WILFREDO QUEZADA PULIDO

Departamento de Filosofía Universidad
de Santiago de Chile
Santiago de Chile, Chile
wilfredo.quezada@usach.cl

<https://orcid.org/0009-0001-0291-686X>

LUIS PAVEZ

Departamento de Filosofía Universidad
de Santiago de Chile
Santiago de Chile, Chile
pavezfloresluis@gmail.com

<https://orcid.org/0009-0000-8478-8277>



RESUMEN

Un desafío clásico para cualquier teoría de causalidad física basada en cantidades conservadas y formuladas en el marco de la teoría de la relatividad especial ha sido distinguir dos regiones, la de las relaciones causales genuinas y aquella de las que son espurias o no causales. En la teoría de procesos causales de Phil Dowe, esto se hace introduciendo un criterio basado en la posesión de una cantidad conservada, el que

* Este artículo se debe citar: Quezada Pulido, Wilfredo & Pavez, Luis. "Entidades físicas, identidad temporal y basura espaciotemporal. Una reconstrucción crítica de la ontología física de Dowe". *Revista Colombiana de Filosofía de la Ciencia* 24.48 (2024): 195-231. <https://doi.org/10.18270/rcfc.4303>

parecería deslindar de manera eficiente entre procesos causales y pseudoprocesos causales. Sin embargo, ante críticas sobre la insuficiencia de su criterio, Dowe ha propuesto modificar su ontología física e introducir, junto con los procesos y los pseudoprocesos, dos tipos de *gerrymanders* físicos, tanto temporales como espaciales. Aquí analizamos y criticamos, desde la perspectiva de la relatividad especial, la caracterización de estas cuasientidades y la necesidad teórica de hacer una distinción taxativa entre ellas. Esto implica una modificación fundamental de la ontología de Dowe y plantea otras dudas acerca de su motivación original.

Palabras clave: causalidad física; cantidades conservadas; transmisión de marca; procesos causales; pseudoprocesos; *timewise gerrymander*; *spacewise gerrymander*.

ABSTRACT

A classic challenge for any theory of physical causation based on conserved quantities and formulated within the framework of special relativity theory, has been to distinguish two regions, that of genuine causal relationships and that of those spurious or noncausal ones. In the causal processes theory defended by P. Dowe, this is done by introducing a criterion based on the possession of a conserved quantity, which would seem to efficiently demarcate between causal processes and causal pseudoprocesses. However, faced with criticism about the inadequacy of his criteria, Dowe has proposed modifying his physical ontology and introducing, along with processes and pseudo-processes, two types of physical gerrymander entities, both temporal and spatial. In this article, we analyze and criticize, from the perspective of ERT, Dowe's characterization of these quasi-entities and the theoretical need to make a clear-cut distinction between them. This implies a fundamental modification of Dowe's ontology and raises other doubts about its original motivation.

Keywords: physical causation; conserved quantities; mark transmisión; causal processes; pseudoprocesses; *timewise gerrymander*; *spacewise gerrymander*.

1. INTRODUCCIÓN

Una forma contemporánea de estudiar la causación física que ha ganado creciente interés en la filosofía de la física contemporánea es aquella basada en entender dicha conexión causal en términos de procesos que conservan cantidades físicas. En su desarrollo conceptual han participado, entre otros, filósofos como Russell, Fair, Aronson, Salmon, y su enunciación canónica ha sido formulada en la última década del siglo XX sobre todo por Phil Dowe. En esta última formulación se caracterizan con relativa claridad aquellos procesos que deberían contar como causales desde el punto de vista de la física contemporánea, pero en particular desde el punto de vista de la relatividad especial. Un proceso causal en dicha teoría es una línea de mundo de un objeto que posee una cantidad física conservada y esta es determinada por las leyes de la física prevaleciente. Junto con ello se sigue que, según la relatividad especial, tal proceso debe respetar el límite para la velocidad de cualquier objeto, la velocidad de la luz en el vacío (límite c), es decir, no son admisibles procesos causales superlumínicos. Esta característica plantea un problema bien conocido cuando la relatividad especial se interpreta causalmente debido a que, como hizo notar Reichenbach en la primera mitad del siglo XX, no es difícil mostrar que podemos obtener efectivamente secuencias físicas, compatibles con la relatividad especial, que violan el límite c y que Reichenbach llamó secuencias causales irreales. Más allá de la realidad física de estas secuencias, este hecho implica que cualquier teoría de causalidad física basada en cantidades conservadas y formuladas en el marco de la teoría de la relatividad especial debe distinguir entre procesos causales genuinos y las mencionadas secuencias irreales. La propuesta original de Dowe implementa metafísicamente dicha distinción diferenciando entre dos tipos de entidades: los procesos causales, como ya han sido definidos, y los pseudoprosos causales, esto es, entidades que no poseen cantidades conservadas y pueden violar c , y que capturan entonces las secuencias señaladas por Reichenbach. Esta ontología dual —inicialmente muy

eficiente para tratar casos clásicos de pseudoprocesos— enfrenta presiones mayores cuando se le plantean ejemplos complejos y, sin embargo, físicamente compatibles con la relatividad especial, por ejemplo, entidades mereológicas cuyas partes individuales poseen cantidades y que, para efectos de la relatividad especial, se comportan como gusanos espaciotemporales clásicos. Ya que tales entidades pueden superar claramente el límite c , Dowe —para acomodarlas en su teoría— ha respondido reinterpretándolas y, al hacer eso, se ha visto forzado consecuentemente a modificar su ontología física introduciendo dos tipos de cuasientidades adicionales, que él denomina en inglés *gerrymanders* físicos y que pueden ser de naturaleza temporal o espacial. De esta manera, Dowe obtiene una taxonomía ontológica tetrádica más compleja asociada a su teoría original. Resulta así inevitable preguntarse en general al menos dos cosas: primero, si la reinterpretación física propuesta por Dowe de estas cuasientidades adicionales es la más correcta, y, segundo, si la consecuente inflación de la ontología física que dicha reinterpretación implica está bien motivada. La argumentación que a continuación desarrollamos busca tratar críticamente ambas cuestiones desde un punto de vista físico y, a partir de ahí, extraer ciertas conclusiones que deberían llevar a una revisión de las bases metafísicas del proyecto causalista de Dowe. Con ese propósito, el presente trabajo se articula en cuatro secciones. La primera examina la motivación de Dowe para proponer *gerrymanders* en su filosofía de la causalidad física y, en particular, *gerrymanders* temporales, analizando su naturaleza más allá de la propia caracterización de Dowe. La segunda realiza de la misma manera el análisis de los *gerrymanders* espaciales. La tercera examina la relación entre ambas cuasientidades y sus semejanzas y diferencias, pero representando estas últimas desde la teoría de la relatividad especial. Finalmente, la última sección elabora las principales conclusiones de nuestro estudio y discute varias de sus implicaciones para la teoría ontológica de Dowe, en particular para su tesis de la identidad temporal de los procesos físicos.

2. *TIMWISE GERRYMANDERS*

En *Physical Causation* (Dowe 2000 cap. 5), así como en otros trabajos (1995a, 1995b), Phil Dowe, en el curso de elaborar una respuesta sistemática a críticas formuladas por W. Salmon en los años noventa (Salmon 1994, 1997, 1998) a su propuesta de causalismo físico basada en la posesión de cantidades conservadas, introduce a la discusión física objetos mereológicos especiales con los que, de alguna manera, ya nos había familiarizado Goodman con su propuesta de objetos verdes (objetos que han sido observados como verdes hasta t o que serán azules después de t), en el contexto de desafiar nuestras concepciones filosóficas comunes acerca de inducción. Estas cuasientidades reciben el nombre idiosincrático en inglés de *gerrymanders*, una expresión difícilmente traducible al español. Dichas cuasientidades pueden tener naturaleza temporal y, en ese caso, se denominan en inglés *timewise gerrymanders* (de aquí en adelante TWG), o naturaleza espacial y en tal caso se denominan en inglés *spacewise gerrymanders* (de aquí en adelante SWG).

En su respuesta a Salmon, Dowe presupone que hay una distinción fundamental entre una secuencia continua de parches iluminados en la pared por un chorro de luz y los puntos de luz que aparentan moverse en ella. Según Dowe, los parches en la secuencia son los que poseen una cantidad física y , por ende, constituyen legítimos procesos causales (PC de aquí en adelante) que no rebasan la velocidad de la luz. En cambio, los puntos de luz móviles no poseen energía y aparentan moverse por sobre dicha velocidad dando origen a lo que él denomina pseudoprosos (PSP de aquí en adelante). La réplica de Salmon a esta primera observación de Dowe es que, si los puntos de luz no poseen cantidades conservadas (CC de aquí en adelante) y sí las partes iluminadas de la pared, entonces una línea de mundo de las partes de la pared representa a un objeto que posee CC en la dirección del movimiento y, sin embargo, no transmite dichas cantidades en esa misma dirección. En relación con esto, Dowe dice:

Salmon (1994 308) da un ingenioso contraejemplo a esta respuesta, pidiéndonos que consideremos “la línea de mundo de la parte de la superficie de la pared que está absorbiendo energía como resultado de ser iluminada” (1994 308). Este objeto “*gerrymandered*” es el conjunto de todos los parches de la pared que se iluminan secuencialmente, tomados solo por el tiempo en que están siendo iluminados. Salmon argumenta que este objeto posee energía durante el intervalo relevante, pero no transmite energía. La implicación es que la línea de mundo de este objeto no es un proceso causal, y sin embargo el objeto posee energía; por lo tanto, necesitamos invocar la noción de transmisión —la posesión no es suficiente (Dowe 2000 98).

Dowe se muestra de acuerdo con Salmon, como ya se dijo, en que son los parches de la pared los que poseen CC al ser iluminados por los puntos de luz, pero que aun así esto no es suficiente para ser calificados como un PC en el sentido en que su teoría causal de cantidades conservadas define estos últimos¹. Así lo aclara en el siguiente pasaje:

Así, volviendo al ejemplo de Salmon del agregado de las partes de la pared iluminadas secuencialmente, podemos ver que, aunque se genera mediante una única descripción, implica cierta uniformidad y muestra una continuidad espacio-temporal, se trata sin embargo de un *timewise gerrymander*, y no de un objeto en mi definición, ya que no muestra identidad en el tiempo. La mancha en sí es un objeto (aunque no es causal), y la parte completa de la pared es un objeto (como las diez bolas de billar), pero el *timewise gerrymander* no lo es. Por lo tanto, no es un proceso de ningún

¹ Dowe (2000 90) provee la siguiente caracterización clásica de su teoría causal basada en cantidades conservadas, sintetizada mínimamente en dos premisas:

CC1: un *proceso causal* es una línea de mundo de un objeto que posee una cantidad conservada.

CC2: una *interacción causal* es una intersección de líneas de mundo que envuelve el intercambio de una cantidad conservada.

Hitchcock (cf. 2004) ha refinado esta teoría concentrándose en las interacciones causales.

tipo, y mucho menos un proceso causal; en mi opinión, es basura espaciotemporal (Dowe 2000 101).

La discusión con Salmon, como se recordará, reside en la cuestión acerca de qué criterio resulta ser el más adecuado para distinguir procesos causales de procesos no causales o pseudoprocesos en el interior de la teoría de la relatividad especial (TRE de aquí en adelante), un desafío planteado en los años veinte por Hans Reichenbach (1958 147-149). Salmon propuso originalmente el criterio de *transmisión de marca* como lo que caracteriza la capacidad de un proceso causal de transferir una cantidad física. En cambio, Dowe propuso el criterio solo de *posesión* de dicha cantidad como el criterio de individuación que caracteriza un proceso como un proceso causal. Frente al contraejemplo de Salmon, que parece violar el criterio de posesión, Dowe, como se explicó, apela a la distinción entre TWG y SWG, y, como resultado, califica dicho contraejemplo no como un PC o un PSP, sino simplemente como un TWG, es decir, una cuasientidad que posee circunstancialmente cantidades físicas, pero no es parte del dominio de la física teórica. Por ello, Dowe rápidamente termina considerándola “basura espaciotemporal” (*spatiotemporal junk*), puesto que, según su argumentación, no podría calificar como un proceso en ninguna teoría que respete los preceptos de la relatividad especial². Un rasgo ontológico esencial a los PC, los PSP y los SWG (y, por cierto, a cualquier entidad que una teoría física admita como tal), y que Dowe toma como primitivo en su teoría, es que solo ellas pueden gozar de identidad a través del tiempo (Dowe 2000 88-91, 107-109). Los TWG parecen efectivamente carecer de dicho rasgo. Esto explica, en general, por qué Dowe los califica de entrada como basura física, y será una de las cuestiones que estaremos discutiendo en este trabajo. Sin embargo, antes de detallar las razones específicas por

² Véase Choi (2002) para un análisis crítico de los contraejemplos de Salmon, basado también en el concepto de TWG pero distinto del ofrecido por Dowe. El análisis de Choi no examina ni cuestiona la distinción misma TWG-SWG propuesta por Dowe, que es justamente el foco del presente artículo.

las cuales él es conducido a concluir esto, veamos primero cómo Dowe intenta caracterizar este tipo de objetos en el contexto de su propia teoría de causalidad física. Parte afirmando lo siguiente:

Según la teoría de CC s, hay procesos causales tales como bolas de billar que ruedan a través del paño, y pseudo procesos tales como sombras y puntos de luz. ¿Es esto exhaustivo de todos los elementos que se puedan representar como ocupando una región del espacio-tiempo? La respuesta es no; hay también “basura espaciotemporal” —artefactos que no son procesos del todo en la definición de CC s. Un ejemplo es lo que anteriormente llamé los *timewise gerrymander*. Un *timewise gerrymander* es un supuesto objeto definido sobre un intervalo del tiempo donde la definición cambia en un cierto plazo (el objeto supuesto es realmente objetos diferentes en diversos momentos) (Dowe 2000 99).

Dada esta cita, y según todo lo ya dicho, podemos inferir las siguientes dos propiedades generales de un TWG:

1. Los TWG son extensiones temporales compuestas de diferentes procesos temporales parciales, donde cada uno de ellos representa un objeto genuino diferenciable de los otros en el transcurso del tiempo y, por ende, nunca se configura como un objeto único. Esta es entonces la *propiedad mereológica temporal o de temporalidad de las partes* de un TWG.
2. De 1) se sigue que los TWG carecen, como un todo, de identidad objetiva en el tiempo, una propiedad que poseen los procesos, sean o no causales, según Dowe. Esta es la *propiedad de la ausencia de identidad integral en el tiempo* de un TWG.

Como vimos, esto llevará a Dowe a concluir que los TWG no son ni procesos ni, en general, entidades físicas. Es decir, Dowe considera que 1 y 2 son suficientes para afirmar la siguiente tesis:

Los TWG no pertenecen a una ontología (y epistemología) física que admita CC pues podemos distinguirlos claramente de los PC y de los PSP y de cualquier entidad que sea parte de aquella y que sea contemplada en el trabajo efectivo que realizan los físicos. De ahí también que se deban excluir como simple basura espaciotemporal.

Por contraste, en una teoría causal de CC que presupone TRE, según Dowe, tanto los PC como los PSP son líneas de mundo que representan procesos y pueblan el espacio-tiempo de Minkowski y, por ello, son huéspedes ontológicos legítimos en una teoría de ese tipo. Dowe entonces procede a ejemplificar algunos TWG con el fin de contrastarlos con PC. He aquí un primer ejemplo:

Un ejemplo de un *timewise gerrymander* es el objeto supuesto X definido como:

- para $t_1 \leq t < t_2$, X es la moneda en mi bolsillo
- para $t_2 \leq t < t_3$, X es el lápiz rojo en el escritorio
- para $t_3 \leq t < t_4$, X es mi reloj

Obsérvese que X ocupa una región determinada del espacio-tiempo, y que en cualquier momento del intervalo Δt , X posee CCs como momentum (aunque no en sentido estricto, porque algo debe ser un objeto para poseer una cantidad conservada). Claramente, hay innumerables *timewise gerrymander* (Dowe 2000 99).

Resulta claro que los elementos constituyentes del TWG ejemplificado exhiben distintas posiciones espaciales en los diferentes tiempos. Además, resulta igualmente claro por qué este tipo de TWG no debe ser tomado como un

proceso, ya que difícilmente objetos diferentes y sucesivos en el tiempo, que se hallen en distintas partes del espacio, podrían ser considerados parte de un único objeto. Un TWG de este tipo no calificaría ni como proceso de sentido común ni como proceso físico, y menos aún como proceso causal. Parece físicamente implausible imaginar la sucesión temporal de una moneda, un lápiz y un reloj, en distintos lugares del espacio, como un objeto idéntico en el tiempo (incluso si los distintos lugares estuvieran muy próximos entre sí y ubicados en una misma línea espacial).

Según lo anterior, podemos concluir con Dowe que la suma mereológica de distintos objetos o procesos, separados espacialmente, y tomados en tiempos continuos y sucesivos, no es un proceso o una entidad física de ningún tipo. La figura 1 muestra el mencionado TWG en el espacio de Minkowski en dos dimensiones.

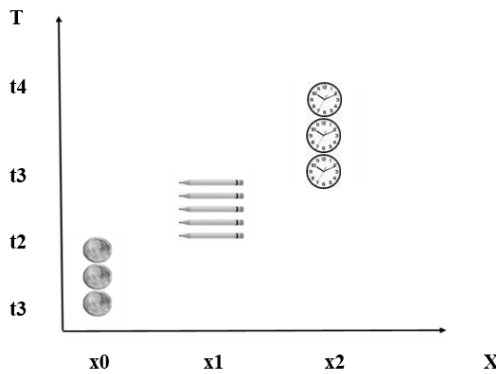


Figura 1. Representación de un TWG en un espacio tiempo bidimensional de Minkowski, donde los objetos constituyentes están en distintos lugares en distintos tiempos sucesivos. Cada objeto podría ser representado por una línea vertical continua, sin embargo, se ha dibujado cada objeto individualmente con fines ilustrativos.

Fuente: elaboración propia.

Consideremos ahora un segundo ejemplo de TWG dado por Dowe y representémoslo de nuevo en un espacio-tiempo bidimensional de Minkowski. Él lo define de la siguiente manera:

En un juego de billar, X será la bola más cercana a la bola negra en el siguiente sentido:

- para $t_1 \leq t < t_2$, X es la bola rosada
- para $t_2 \leq t < t_3$, X es la bola roja
- para $t_3 \leq t < t_4$, X es la bola blanca

Aquí X es un *timewise gerrymander*, ocupando una región del espacio-tiempo y poseyendo cantidades conservadas. (Dowe 2000 100)

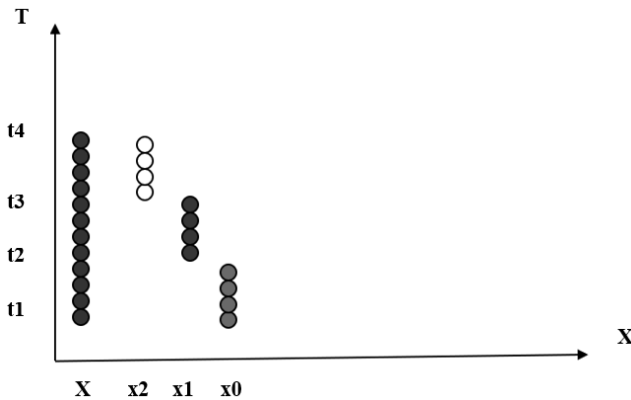


Figura 2. Representación de un TWG en un espacio-tiempo de Minkowski en dos dimensiones, en el cual las bolas que están en las posiciones X_0 , X_1 y X_2 van apareciendo a medida que pasa el tiempo cada vez más cerca de la bola negra, que está en la posición X .

Fuente: elaboración propia.

Este TWG se diferencia del anterior en que sus constituyentes son objetos físicamente muy similares entre sí, distinguiéndolos solo los colores. El objeto sombreado más a la izquierda es la bola negra tomada como referencia en el TWG en la cita más arriba y está compuesto por las tres bolas más a la derecha. El TWG entonces como objeto mereológico posee CC (cantidad de movimiento de cada bola que se acerca). La línea de mundo de este TWG está compuesta de tres líneas verticales sucesivas separadas espacialmente, y como cada componente es un objeto distinto, el TWG completo no puede clasificarse como un objeto con identidad en el tiempo, de manera que la secuencia temporal de distintos objetos o PC similares en la figura 2, aunque representada en una cierta región del espacio, no constituye un proceso. Los ejemplos anteriores, no obstante, plantean algo crítico pues el TWG, como dijimos, posee efectivamente CC en cada región del espacio-tiempo en la cual es descrito, como concede Dowe, y según la teoría causal de CC, esta parece ser, de hecho, la propiedad distinguidora de un proceso causal. Justamente el no poder cumplir con este criterio es el que hace que las sombras, por ejemplo, no puedan contar como causales sino solo como pseudoprocesuales. Dowe intenta aclarar esta cuestión mediante el siguiente nuevo ejemplo de TWG:

Los *timewise gerrymanders* exhiben a veces consistencia de una cierta característica. Por ejemplo, en una caja de moléculas, tomemos a X como cualquier molécula con momento p_x , justo durante el tiempo que tiene dicho momento. X es otra vez un *timewise gerrymander*, ocupando una región del espacio-tiempo y “poseyendo” cantidades conservadas, pero aquí X muestra un momento estable. El objeto tiene consistencia de una cierta característica sobre su historia entera (Dowe 2000 100).

Este ejemplo de TWG muestra una CC estable durante todo el tiempo de su definición en la medida que X en la caja de moléculas posee a través de su historia una cantidad de momento dada en un tiempo específico. Dicha posesión es, podemos decir, la posesión mereológicamente estable de una cantidad conservada. Pese a esto, sin embargo, ese hecho no hará contar a X como un PC o un proceso

en general pues nuevamente el factor determinante para descartarlo como proceso es su falta de identidad en el tiempo. Así, tenemos una tercera propiedad que agregar a la caracterización de un TWG:

3. Un TWG puede poseer una cantidad conservada en relación con sus componentes mereológicos y dicha posesión puede ser estable a lo largo de su trayectoria en el espacio-tiempo. Esta es la propiedad de la *posesión mereológicamente estable de una CC en un TWG*.

La propiedad 3 indica que, a diferencia de los PSP, un TWG puede exhibir información física de manera estable y persistente. Aun con esto, es claro también que la cantidad poseída no se transmite o transfiere al resto de los componentes del TWG³. Esta última característica puede ser considerada otra propiedad esencial de los TWG que, aunque no es destacada por Dowe, convendrá distinguir aquí para entender mejor la argumentación de este último y, a su vez, será funcional a nuestros propósitos argumentativos. Es el siguiente:

4. Los componentes de un TWG, aunque exhiban individualmente CC, no las transfieren o transmiten entre sí. Esta es la propiedad de *la intransferibilidad física de CC* de los TWG.

El próximo ejemplo seleccionado resulta muy conveniente para el objetivo de nuestro trabajo, pues revela que los TWG no solo pueden presentar CC parciales

³ Aun cuando todos los componentes de un TWG posean o tengan la misma cantidad conservada (la misma masa y la misma velocidad ($p_x = mv_x$)) dicha cantidad no puede estar en una relación de intercambio de un objeto componente a otro objeto componente del TWG en la dirección espacial en la que aparece el TWG. Como Dowe muestra en su discusión con Salmon, la falta de identidad en un TWG elimina como un todo cualquier tipo de persistencia o permanencia de sus propiedades dinámicas en el tiempo, de ahí que sus partes pueden poseer CC, pero por ser distintos objetos y estar ontológicamente desvinculados entre sí, en la dirección del movimiento aparente, sus CC están desconectadas y no se transfieren de un lugar a otro y menos aún se transmiten.

entre sus componentes, como los casos ya mostrados, sino que además pueden presentar continuidad espaciotemporal, al igual que muchos procesos genuinos en física clásica. De hecho, es el ejemplo final invocado por Dowe para rechazar el contraejemplo de Salmon:

Consideremos una línea de diez bolas de billar estacionarias y contiguas. Dejemos a X ser la suma mereológica de la primera bola durante el primer intervalo, más la segunda durante un intervalo de tiempo inmediatamente subsiguiente, y así sucesivamente para las diez bolas. Entonces X es un *ti-mewise gerrymander*, representado en un diagrama del espacio-tiempo por una línea diagonal de aproximadamente una bola de billar de ancho (Dowe 2000 100).

La similitud mostrada de este TWG con un proceso, según lo define la teoría de CC, nos permitirá, hacia el final, conectarlo conceptualmente con un PSP. De momento, analicemos la suma mereológica X de la secuencia temporal de las diez bolas. Cada bola es un objeto diferente a su vecina inmediata, aunque similar en su forma y masa. La secuencia X puede ser representada entonces en un espacio-tiempo bidimensional de Minkowski (figura 3).

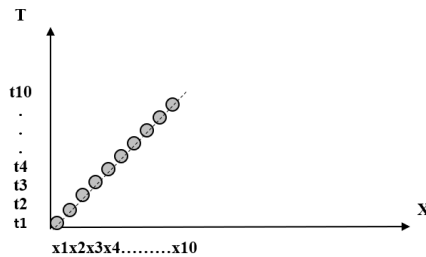


Figura 3. Representación de un TWG en un espacio-tiempo de Minkowski en dos dimensiones, donde cada bola, como parte de una fila de diez bolas, es considerada en forma sucesiva en el tiempo.

Fuente: elaboración propia.

Esta representación permite formular una quinta propiedad que caracteriza a un TWG, dado que sus componentes pueden ser representados ordenadamente en el espacio-tiempo como líneas de tipo tiempo, tipo luz o tipo espacio (Dowe 2000 90-ss)⁴. Podemos formularla de la siguiente manera:

5. Un TWG puede ser representado en el espacio tiempo como una entidad que despliega continuidad en virtud de que cada una de sus partes espaciales y puntuales parecen evolucionar temporalmente en dicho marco. En otras palabras, esta es la propiedad de *continuidad espaciotemporal* de un TWG.

Así, es visualmente atractivo interpretar la figura 3 como describiendo la línea de mundo de un objeto único que se extiende en el tiempo en virtud de las propiedades dinámicas de sus partes (dadas las características de parecido observacional, de propiedades dinámicas y contigüidad de las bolas de billar). Sin embargo, como indica Dowe, sigue sin ser un objeto genuino y no corresponde, por tanto, a un proceso causal ni pseudocausal —y, en general, a ninguna entidad genuinamente física— pese a que sus partes sean similares, cada una de ellas posea CC y representen un todo continuo graficado por una línea de mundo.

De esta manera, vemos ahora que un TWG podría ser formalmente caracterizado por las cinco propiedades que hemos destacado. Aunque es lógicamente posible que alguna propiedad pueda ser derivable de otra de las propiedades, no estaremos interesados aquí en defender en detalle la primitividad lógica o física de cada una de ellas. Sin embargo, resulta claro que físicamente la argumentación

⁴ El hecho de que este TWG pueda ser representado en un diagrama de Minkowski por una línea diagonal nos permite imaginar que dicha línea puede caer dentro del cono de luz, en el cono de luz o fuera de él. Esto dependerá, como es obvio, de la relación espaciotemporal que se genere en cada caso especial de TWG. Por ende, no es ni intuitiva ni formalmente problemático que el TWG aparezca en un diagrama de Minkowski superando la velocidad de la luz. Según esto, el caso graficado en la figura 3 podría perfectamente estar fuera del cono de luz, sin perder su continuidad espaciotemporal. Por otra parte, lo anterior no excluye que la línea del TWG pueda ser una curva.

de Dowe presupone no solo las propiedades 1 y 2, sino también la propiedad 4, es decir, la naturaleza temporal de sus partes y la ausencia de integridad temporal entre ellas, más la intransferibilidad de información física entre estas. Tendremos más que decir sobre esto un poco más adelante.

3. DE LOS *TIMWISE GERRYMANDERS* A LOS *SPACEWISE GERRYMANDERS*

Como ya indicamos, Dowe admite otro tipo de entidad *gerrymandered*: los homólogos espaciales de los TWG, esto es, los *spacewise gerrymanders* o SWG. Lo importante es que para este autor ellos sí pueden participar de la ontología física de una teoría causal de CC. Es decir, Dowe los considera objetos genuinos y completos. Como una forma inicial de ilustrarlos, dice lo siguiente:

Los *timewise gerrymanders* deben ser distinguidos de los *spacewise gerrymanders*. Un ejemplo de *spacewise gerrymander* es el supuesto objeto Y que consiste en mi reloj más la pluma roja en mi escritorio más la moneda en mi bolsillo. [...] la representación del espacio-tiempo de Y consiste en tres líneas verticales que coexistan sobre el intervalo entero. (2000 99)

Como se podrá notar, este ejemplo de un SWG resulta similar al primer ejemplo de la sección anterior de un TWG, pero con la diferencia de que en este caso los objetos componentes coexisten contemporáneamente durante todo el tiempo determinado por la definición. En la figura 4, mostramos las líneas de mundo de este SWG en el espacio de Minkowski.

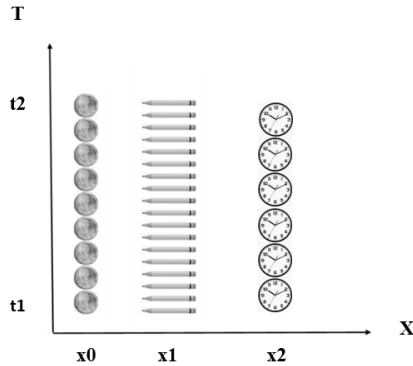


Figura 4. Representación de un SWG en un espacio-tiempo bidimensional de Minkowski, donde tres objetos diferentes separados espacialmente coexisten en reposo mientras transcurre el tiempo

Fuente: elaboración propia.

Se deberían notar dos cosas acerca de este ejemplo. La primera concierne al transcurso temporal: en un SWG es fundamental que los constituyentes coexistan en un periodo o intervalo de tiempo por mínimo que sea, no importando si se encuentran distantes o contiguos. Esta característica será importante en la argumentación que desarrollaremos más adelante. La segunda cosa a notar es que los distintos objetos coexistentes componentes del SWG se encuentran siempre claramente diferenciados espacialmente. Incluso, aunque ellos estuviesen muy cercanos entre sí y de manera continua, sería fácil aún distinguir el uno del otro. La coexistencia de estos tres objetos constituyentes del SWG no resulta suficiente para interpretar a este último como un único objeto. Sin embargo, podemos encontrarnos con algunos casos de SWG donde sus componentes exhiban una gran semejanza entre sí, tanto desde una perspectiva observacional como desde una perspectiva física. Veamos cómo se da esto. Retomemos, en primer lugar, el último TWG de la sección anterior, que consiste en una secuencia X de diez bolas

de billar tomadas una a una y de manera sucesiva en el tiempo, pero esta vez fijemos la atención en la sola coexistencia de sus componentes, es decir, tomemos las diez bolas de billar en su conjunto como un SWG y nombrémoslo con la letra Y. En relación con el TWG, Dowe nos advierte: “[O]bserve que X no debe ser confundido con el objeto Y que consiste en la línea de mundo de diez bolas, que *es un objeto genuino, y que es representada por un bloque vertical de diez bolas de ancho*” (2000 100; el destacado en cursiva y negrita es de Dowe). En otras palabras, para Dowe las diez bolas contiguas en el espacio califican como un objeto único, constituyendo un SWG similar a un cuerpo extenso y rígido, pero hecho de bolas de billar desconectadas entre sí. Lo podemos graficar mediante la figura 5.

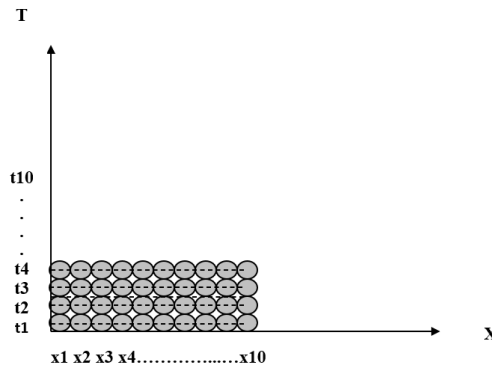


Figura 5. Representación de un agregado de bolas en una dimensión espacial X que evolucionan juntas en el tiempo, generando una línea de simultaneidad en el espacio de Minkowski, indicada por cada línea segmentada paralela al eje X
Fuente: elaboración propia.

Este conjunto continuo, simultáneo y coexistente de bolas de billar, que evoluciona en el tiempo, debe ser, según Dowe, interpretado como un objeto físico y, por ello, similar en todo a un proceso causal genuino (Dowe 2000 91).

Desde el punto de vista de nuestra argumentación, lo fundamental es que las propiedades 1 y 2 de los TWG, en el caso de un SWG, deben ser modificadas a la siguiente versión:

1') Los SWG son extensiones espaciales compuestas de diferentes procesos individuales espaciales, donde cada uno de ellos es un objeto genuino diferenciable del resto pero evolucionando en un mismo transcurso o intervalo de tiempo, por mínimo que sea. Por ende, esta entidad puede evolucionar como un objeto espacialmente extenso en el tiempo y constituye un objeto único. Esta es entonces la *propiedad mereológica espacial* de un SWG.

Dada esta propiedad de los SWG, se sigue inevitablemente, como admite Dowe, la siguiente propiedad:

2') Los SWG poseen identidad objetual en el tiempo, una propiedad que comparten con los procesos causales. Esta es la *propiedad de identidad en el tiempo* de un SWG.

Así, los SWG participan de la ontología física causal por dos razones fundamentales: (1) porque poseen CC (cada bola tiene momentum lineal, por ejemplo, y el todo mereológico hereda la suma de esos momentum)⁵, y (2) porque tienen identidad en el tiempo y ella puede ser efectivamente representada en un diagrama de Minkowski. Sin embargo, para hacernos una idea completa de lo que cuenta como un SWG como distinto a un TWG, debemos examinar adicionalmente las propiedades 3) a 5) que le hemos adjudicado a estos últimos. La versión de 3) para los SWG es coincidente con ella y es la siguiente:

⁵ Hay que hacer notar que en la teoría de CC de Dowe es posible aceptar una cantidad de movimiento o momentum igual a 0 (Cf. 2000 93).

3') Un SWG puede poseer cantidades conservadas en relación con cada uno de sus componentes mereológicos y dicha posesión puede ser estable a lo largo de su trayectoria en el espaciotiempo. Esta es la propiedad de la *posesión mereológicamente estable de una CC en un SWG*.

No obstante esta coincidencia, un SWG no puede coincidir con un TWG en relación con la propiedad 4) pues, al contrario de un TWG, tanto sus componentes como el propio SWG pueden efectivamente exhibir transferibilidad física de sus cantidades, ya sea entre sí como con otros objetos físicos. De esta manera, debemos admitir la siguiente como una de sus propiedades:

4') Las cantidades conservadas en un SWG pueden ser transferidas tanto entre sus componentes como aditivamente por el propio SWG a otros objetos físicos. Esta es la propiedad de *la transferibilidad física de CC de los SWG*.

Finalmente, otra propiedad característicamente distinta de los SWG en relación con los TWG emerge cuando se analiza cómo evoluciona el primero en el continuo espaciotemporal de la TRE. Ahí se observa de inmediato que la evolución en el tiempo de la entidad física creada por el SWG no puede ser modelada como un punto espacial que da origen a una línea espaciotemporal, sino como una línea espacial que evoluciona temporalmente formando una superficie espaciotemporal. De esta manera, obtenemos la última propiedad:

5') Un SWG puede ser representado en el espacio tiempo como una entidad que despliega continuidad en virtud de evolucionar como un todo espacial no puntual generando una superficie espaciotemporal en dicho marco. Esta es la propiedad de *continuidad espaciotemporal extensional de un SWG*.

Entonces, la propiedad 2') junto con las propiedades 4') y 5') representan las características diferenciadoras importantes de los SWG en relación con los TWG. Por otra parte, respecto a los PSP, según la teoría de Dowe, las características diferenciadoras importantes serían 3') y 4') pues aquellos no poseen CC y, por ende, no las transmiten. Sin embargo, ambas entidades concurren en tener identidad en el tiempo y, por tanto, coinciden en la propiedad 2').

Ahora que hemos precisado e ilustrado, desde la mirada del sentido común, los distintos modos en que se pueden presentar los TWG y los SWG en el espacio-tiempo, veremos, a continuación, cómo ellos se pueden relacionar entre sí, examinándolos desde la perspectiva de la TRE.

4. TWG Y SWG: SEMEJANZAS Y DIFERENCIAS EN EL ESPACIO-TIEMPO DE MINKOWSKI

Como ya se dijo, un TWG es una supuesta entidad que se puede representar en el espacio tiempo de Minkowski. Por lo dicho también (nota 4 al pie de página), los TWG no necesitan tener límites de velocidad y pueden presentarse en cualquier región del espacio-tiempo. De esta manera, si nos focalizamos en un TWG superlumínico y lo graficamos en el espacio de Minkowski, aparecerá representado por una línea de mundo de tipo espacio, lo que indica no solo una relación entre espacio y tiempo mayor que c (o velocidad mayor que c), sino que, además, su orden temporal se verá alterado. Esto último es coherente con el diagnóstico de la teoría de CC de Dowe puesto que se trataría de la representación de basura espaciotemporal.

Por otro lado, un SWG —considerado ahora un objeto extenso genuino— no podría alcanzar velocidades ni lumínicas ni súper lumínicas ya que, según la TRE, los objetos materiales no pueden alcanzar la velocidad de la luz, menos aún superarla. De esta manera, el orden temporal de un SWG deberá permanecer inalterado en su evolución en el espacio-tiempo. Desde la perspectiva

de la TRE, entonces, esto último lo distinguirá fundamentalmente de un TWG. Esta forma de caracterizar la diferencia entre las dos entidades *gerrymandered* tiene consecuencias fundamentales que pasamos ahora a explorar.

En primer lugar, si retomamos el examen de un TWG, por ejemplo, el caso de las bolas de billar sucesivas en la figura 3, podemos admitir, como hemos dicho, la posibilidad física de relacionar el espacio cubierto por ellas y el tiempo transcurrido, de tal manera que el TWG pueda resultar ser superlumínico. Esto significa, según la TRE, que es posible contar con otro sistema de referencia, con velocidad relativa al primero, donde se observen todos los componentes del TWG al mismo tiempo o de manera simultánea. Es decir, el TWG súper lumínico completo para un primer observador, en todo el tiempo de su definición (Δt), se transforma en el inicio temporal simultáneo de un SWG para un segundo observador (ya que en este último sistema de referencia se consideran todas las bolas en simultáneo en el instante t_0). Además, es completamente posible imaginar un tercer observador en otro sistema de referencia, en movimiento relativo a los otros dos sistemas, para el cual el TWG original invierte su orden temporal ($-\Delta t$).

De manera recíproca, si se considera ahora un sistema de referencia, donde en t_0 (tiempo instantáneo) se dé el inicio simultáneo de un SWG (como el bloque de las diez bolas de billar tomadas todas juntas, por ejemplo), entonces se podrá contar con algún sistema de referencia alternativo, con velocidad relativa al primero, que observe sucesivamente cada una de las bolas, en un sentido del tiempo (Δt), mientras que otro sistema alternativo, con velocidad relativa, observaría el tiempo sucesivo en sentido inverso ($-\Delta t$).

En consecuencia estricta con nuestro análisis, se sigue que un instante de tiempo t_0 del tiempo de definición de un SWG, para un dado observador, será un TWG para otro observador (Δt). Así, cada instante de tiempo del SWG, en un sistema de referencia, tendrá su correspondiente TWG en otro sistema, y recíprocamente. Luego, un SWG que se sostiene, según Dowe, como objeto extenso con

identidad en el tiempo debido a la coexistencia instantánea de los distintos componentes durante cada uno de los instantes de tiempo del intervalo de definición, relativiza su identidad pues se vuelve en otro sistema, para cada instante de dicho intervalo de tiempo, un TWG y, por lo tanto, un conjunto de TWG en el intervalo de tiempo de definición del SWG. Es decir, un objeto extenso y genuino en la teoría de CC de Dowe, con identidad en el tiempo, se transformaría en cada instante de tiempo (t_0, t_1, t_2, \dots) —usando la expresión de Dowe— en basura espatiotemporal para un segundo sistema de referencia. Como sabemos, una consecuencia física de la TRE es que la simultaneidad espacialmente puntual resulta ser absoluta y que la simultaneidad espacialmente extensa, en la dirección del movimiento relativo de los sistemas, resulta ser relativa. Obviamente el argumento que hemos desarrollado hasta aquí sobre la relación entre un SWG y un TWG concierne directamente a la noción de simultaneidad relativa⁶.

A continuación ilustramos con dos ejemplos, en el espacio-tiempo de Minkowski, la argumentación recién expuesta. Comenzaremos con un TWG sublumínico, fenómeno que mantiene el orden temporal en todos los sistemas de referencia, para pasar luego a un TWG superlumínico, que pierde el orden temporal y se relaciona directamente con los SWG.

Tomemos una serie de cinco monedas de cincuenta pesos chilenos contiguas entre sí y en una fila en el eje horizontal, como muestra la figura 6.

⁶ Es importante distinguir entre la simultaneidad absoluta y la relativa, pues es justamente dicha distinción la que generará discusiones adicionales sobre la naturaleza del tiempo en TRE como, por ejemplo, aquella relativa a la transitividad del ahora real, al privilegiar la simultaneidad espacialmente extensa (cf. Sklar 1992 cap. 2), o aquella concerniente a la temporalización del espacio, a través del orden lineal y cónico (véase Robb 1914 para un tratamiento pionero de esta cuestión, en el que se privilegia la simultaneidad espacialmente puntual; para un análisis útilmente detallado de la concepción geométrica de Robb, véase Torretti 1983).



Figura 6. Monedas de 50 pesos chilenos.

Fuente: elaboración propia.

Elijamos una relación espaciotemporal entre sus respectivos lugares de izquierda a derecha y en tiempos sucesivos, de tal manera que dicha relación sea menor que c . Con ello se obtiene, para un observador en reposo, un TWG sublumínico, como se aprecia en la figura 7.

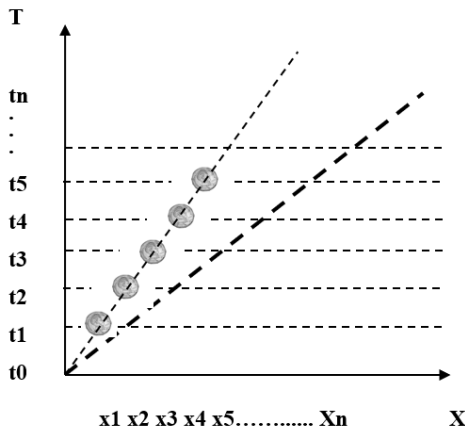


Figura 7. Cada línea segmentada horizontal y paralela al eje X es una línea de simultaneidad. La línea diagonal, segmentada y gruesa, representa el límite de la velocidad de la luz.

Fuente: elaboración propia.

Ahora, si al esquema anterior agregamos otro observador, representando otro sistema de referencia que se mueve hacia la derecha respecto del primero, digamos S' (con coordenadas X' y T'), se obtiene el esquema en el espacio-tiempo de Minkowski que se muestra en la figura 8.

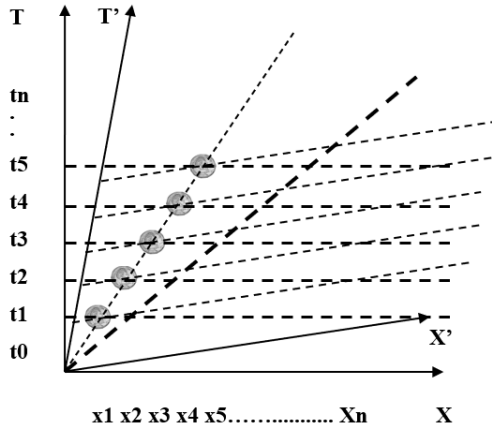


Figura 8. Cada línea segmentada horizontal paralela al eje X es una línea de simultaneidad en S y cada línea segmentada paralela a X' es una línea de simultaneidad del sistema S' . La línea diagonal, segmentada y gruesa, representa el límite de la velocidad de la luz. Fuente: elaboración propia.

La figura 8 ilustra que tanto para S como para S' el TWG mantiene el orden temporal. Más aún, si agregamos otros sistemas de referencia, ya sea que se muevan hacia la derecha o hacia la izquierda con respecto al sistema original S , el orden del tiempo se mantendrá inquebrantable para todos ellos.

Consideramos ahora un segundo TWG, con las mismas monedas, pero esta vez superlumínico (es decir, elijamos una relación espacio-temporal cuya línea de mundo quede fuera de la frontera del cono de luz, representada por la

línea gruesa segmentada). En este caso, los dos observadores mencionados, S y S', mantendrán aún el orden temporal y sucesivo del TWG superlumínico, como mostramos en la figura 9.

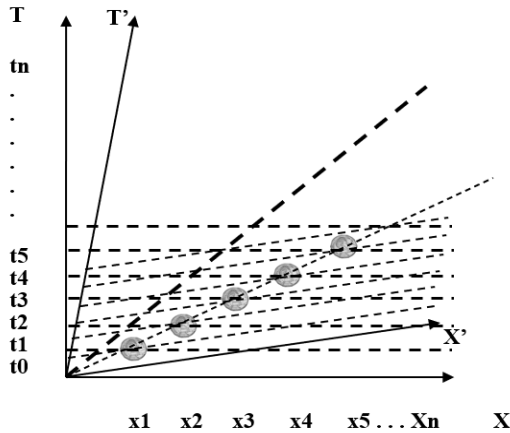


Figura 9. Representación de una serie de monedas, tanto para un observador S como para uno S', que forman un TWG bajo una relación espaciotemporal superlumínica en el espacio de Minkowski. Cada línea segmentada paralela al eje X es una línea de simultaneidad en S y cada línea segmentada paralela a X' es una línea de simultaneidad del sistema S'. La línea diagonal gruesa y segmentada representa el límite de la velocidad de la luz

Fuente: elaboración propia.

Nótese que, aunque el TWG es superlumínico, todavía disponemos de observadores en movimiento relativo para quienes el orden temporal no se vea alterado. Sin embargo, puede darse el caso de que un tercer observador alternativo, digamos un sistema S'', cuyos ejes serán X'' y T'', moviéndose también hacia la derecha, según S, observe las partes del TWG simultáneamente, lo que evidenciaría su conversión en una línea de simultaneidad correspondiente al inicio temporal de un SWG (como una coexistencia simultánea en t_0). En la figura 10

ilustramos esta situación, donde la parte izquierda de la imagen describe lo recién dicho y la parte derecha su símil simétrico inverso desde la mirada de S'' .

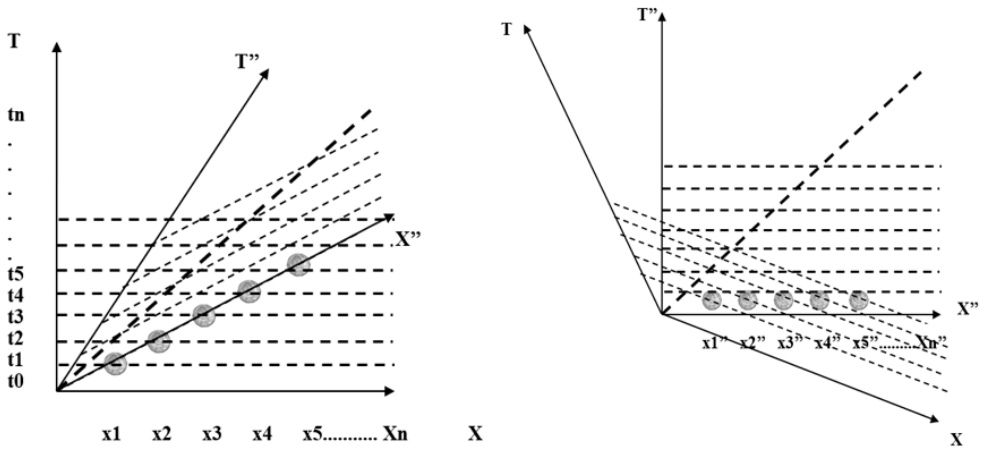


Figura 10. A la izquierda se representa un TWG de la secuencia de monedas con rapidez superlumínica en el sistema S , que se muestra simultáneo en el sistema S'' , el cual se mueve respecto de S . A la derecha, ocurre la misma situación pero en una representación de movimiento relativo inverso entre los sistemas S y S'' , donde para S'' , representado en reposo, es S quien se mueve hacia la izquierda. Cada línea segmentada paralela al eje X es una línea de simultaneidad en S y cada línea segmentada paralela a X'' es una línea de simultaneidad del sistema S''

Fuente: elaboración propia.

Ahora, dado este ejemplo, es fácil imaginar que si se elige otro sistema de referencia (digamos S'''), también moviéndose a la derecha respecto de S , pero esta vez más rápido que el sistema S'' , se observarán las monedas del TWG

con un orden temporal invertido en S''' en relación con el orden temporal atribuido por el observador original del sistema S . La figura 11 muestra esta situación en el sistema S''' , destacándose con líneas gruesas los ejes de espacio y tiempo de dicho sistema, y con líneas gruesas, pero segmentadas, las líneas de simultaneidad.

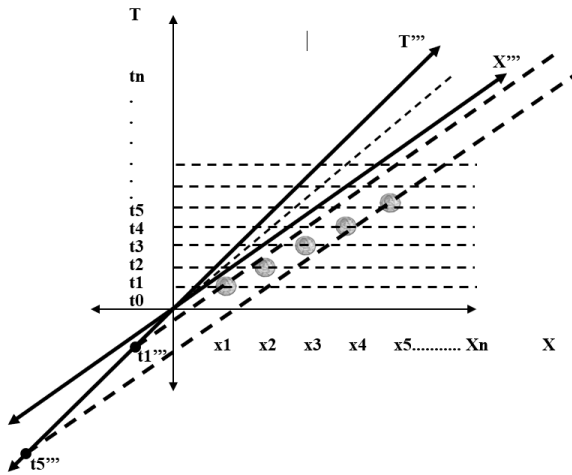


Figura 11. Un observador, S''' , que describe el orden del tiempo de manera invertida respecto de S . En S''' el tiempo 5 ocurre en primer lugar y el tiempo 1 en último lugar. **Fuente:** elaboración propia.

De este modo, lo que se ha mostrado en esta sección es que estamos siempre en condiciones de interpretar una componente contemporánea (simultánea) de un SWG dado como un TWG cuando se relacionan mediante distintos sistemas de referencia, en acuerdo con la TRE y tal como los caracteriza Dowe mediante las propiedades específicas que se les deben adjudicar. Sin embargo, esto evidencia que la propiedad 2') de identidad en el tiempo, solo propia, según Dowe, de

un SWG y no de un TWG, finalmente tiene una componente espacialmente extensa y contemporánea (simultánea) que es relativa a dichos sistemas y que esta relativización altera directamente la pretendida identidad de cualquier SWG. Evidentemente esta interpretación asociada a los sistemas de referencia implica, a lo menos, una vinculación de origen o genealógica entre los TWG superlumínicos y las componentes instantáneas de un SWG y, de ser esto así, muestra inevitablemente que las propiedades 1) y 1') no son intrínsecas a las entidades *gerrymandered*, pues pueden ser adjudicadas a ellas dependiendo solo de los sistemas de referencia que se estén considerando⁷.

5. CONCLUSIÓN, DISCUSIÓN Y PROSPECTOS

Lo argumentado en la anterior sección aporta, a nuestro juicio, evidencia suficiente para plantear una modificación radical de la ontología física propuesta por Dowe para las teorías que suponen que las CC expresan el vínculo causal. La consecuencia radical es que, eliminada la distinción esencial entre TWG y SWG, ahora tendríamos, por un lado, la clase de los procesos físicos —PC y PSP— y, por otro (en la terminología de Dowe), solo basura espaciotemporal, esto es, la clase de las cuasientidades no procesuales, pero, en este último caso, sin identidad temporal físicamente esencial. No existen más cuasientidades con identidad que, por ello, sean objetivables físicamente en el espacio tiempo, esto es, SWG, como lo requería Dowe⁸. Así, en esta nueva ontología contaríamos con dos tipos de

⁷ Una cuestión más general y tal vez más interesante es explorar si otras de las propiedades individualizadas en este trabajo tanto de los SWG como de los TWG resultan ser invariantes o relativas respecto a los sistemas de referencia. Ya que el objetivo primario aquí era analizar la ontología procesual básica de Dowe, dicha exploración puede quedar para un siguiente trabajo. Agradecemos en todo caso esta observación a uno de los pares evaluadores.

⁸ Nótese que la identidad en el tiempo de un SWG —aquello que le otorga precisamente su importe ontológico— es la identidad de una entidad espacialmente extensa. Pero es justamente el tiempo instantáneo de la coexistencia espacial de una parte de ese SWG —dado que la simultaneidad

entidades con identidad temporal —diferiéndose entre ellas en la posesión de CC— y solo un tipo de cuasientidad mereológica que resulta ser superlumínica o no dependiendo de las relaciones espaciotemporales con las que se presenta.

Sin embargo, pese a esta conclusión y la evidencia suficiente que la respalda, es posible avizorar al menos dos vías para replicar a ella: una filosófica y otra física. La primera plantea que nuestra conclusión implica una indistinción o una indiferenciación espuria entre un TWG y un SWG, esto es, la defensa de una tesis según la cual todo SWG es esencialmente un TWG y viceversa. Sin embargo, se puede sostener, existe un rasgo irreducible entre ambos que garantiza su diferencia y que, de paso, apoya la tesis de Dowe de que los objetos o procesos físicos —incluyendo los SWG— se caracterizan por la propiedad de la identidad temporal. Este rasgo se asocia a la caracterización del tiempo que presuponen tanto los TWG como los SWG. En el primer caso, estamos obligados a asumir que un TWG implica una sucesión de eventos en el tiempo y, por ende, un intervalo en un continuo temporal. En cambio, en el caso de un SWG parece evidente que, analizado en cada una de sus componentes temporales, debemos asumir la instantaneidad del tiempo, es decir un tiempo t instantáneo que se deriva de la relación de simultaneidad en la que, en su evolución temporal, se encuentran las partes del todo mereológico. De esta manera, resultaría imposible obtener la indistinción propuesta pues, en el fondo, estamos comparando entidades incomparables, una entidad cuya caracterización implica su proyección en un tiempo continuo con otra entidad cuya caracterización requiere asumir un tiempo instantáneo. Frente a dicha objeción, es importante tener claro lo siguiente:

1. No sostenemos una indistinción absoluta entre SWG y TWG ni tampoco una reducción ontológica de los primeros a los segundos. Solo

extensa es relativa en TRE — lo que hace que, en dicha teoría, esta entidad *gerrymandered* se proyecte como sucesiva en otros sistemas, volviendo superflua su temporalidad como propiedad esencial de su identidad.

hemos estado argumentando que cada componente temporal singular del SWG es transformable, en otro sistema de referencia, en un TWG superlumínico, pero no hemos sostenido que la evolución en el tiempo del SWG sea transformable, esto es, que la extensión temporal total del SWG lo sea. En otras palabras, sostenemos una indistinción entre un SWG y un TWG solo en el caso de esa componente temporal singular.

2. Por otro lado, lo anterior revela un problema más profundo en la propia objeción, que emerge cuando se afirma que, ya que la componente temporal singular del SWG es instantánea, no podría aceptarse su reducción a una entidad proyectada en un tiempo continuo. Lo primero a decir en este caso es que todo SWG, tal como lo caracteriza Dowe, es temporalmente extenso pues fluye en la línea del tiempo, como lo hace un TWG. Esto significa entonces que deberíamos considerar esa extensión temporal parte de un intervalo continuo. Pero, en segundo lugar, ya que un tiempo continuo debe estar compuesto a su vez de tiempos continuos y no de tiempos discretos instantáneos, es claro que una componente singular del tiempo continuo seguirá siendo continua aunque infinitesimal.
3. En otras palabras, al menos en TRE, la contemporaneidad de las partes de un SWG en un tiempo t no implica instantaneidad absoluta, pues este último tiempo sigue perteneciendo al continuo e implica, por tanto, un intervalo, por pequeño que sea⁹.

⁹ Esta consecuencia, natural en TRE, es asumida directamente por el mismo Dowe cuando propone, como ejemplo de un objeto físico, imaginar una vara de mil millas de largo que viene a la existencia en un milisegundo (cf. 2000 90). Podemos imaginar, a su vez, un SWG muy semejante al tipo de entidad propuesta por Dowe que existiera en un milisegundo o en el tiempo infinitamente pequeño que quisiéramos y sus partes —reales o imaginadas— seguirían siendo contemporáneas sin tener que asumir un instante absoluto para ellas. Al asumir la perspectiva de un tiempo diferencial se asume también que interpretamos la naturaleza del intervalo desde una perspectiva ontológica; sin embargo, en TRE lo usual es asumir un punto de vista metodológico, y, desde esa perspectiva, interpretar dicho intervalo o diferencial como un punto temporal o instante.

Creemos que lo anterior es suficiente para descartar la primera objeción. La segunda vía para objetar la propuesta formulada es eminentemente física y apunta más bien a su generalidad. Consiste en hacer ver que si fuese efectiva la realidad de la conexión genealógica de los TWG superlumínicos y las componentes temporales singulares de los SWG, esto debería reflejarse de una manera robusta y penetrante en todo *gerrymander* físicamente imaginable y no solo en aquellos de naturaleza fundamentalmente mecánica que hemos presentado y examinado. Así, se esperaría que se pudiese mostrar dicha conexión en áreas que son un foco preeminente de análisis en la teoría causal de CC de Dowe, por ejemplo, el electromagnetismo, los fenómenos ópticos y la mecánica cuántica. A esto debemos contestar que el desarrollo de aplicaciones más generales de este tipo excedería el alcance acotado de este artículo, sin embargo, una parte importante de dicho trabajo ha sido ya avanzado por los autores en diversas presentaciones y manuscritos, y confiamos que aparecerá en otro lugar (cf. Quezada & Pavez 2021, 2022a, 2022b).

Finalmente, es inevitablemente tentador preguntarse —en el caso de que los resultados y conclusiones anteriores se sostengan—, si una ontología procesual necesita hospedar tres entidades básicas como las señaladas: procesos casuales, pseudoprocesos y entidades *gerrymandered*, en particular, si la física basada en CC requiere mantener el contraste entre procesos y pseudoprocesos causales. Dowe, como sabemos, argumenta persistentemente que dicho contraste es fundamental si queremos contestar al desafío de Reichenbach en TRE y, como se indicó en la primera sección de este artículo, siempre y cuando el contraste se construya, en lo fundamental, apelando a la idea de posesión de cantidades y al concepto de TWG (y, con él, al de SWG) pues, de otra manera, no se podría contestar a la crítica de Salmon. De no hacerlo, la teoría recaería en el criterio, según Dowe, filosóficamente insalvable que Salmon propone para formular el contraste entre ambos, a saber, el criterio de transmisión de marca. Ante esta cuestión, lo que podemos decir es que nos parece que ahora hay cierta base para replantearse

el problema de la naturaleza ontológica intrínseca de los PSP. En la descripción de Dowe, un PSP sublumínico (por ejemplo, la traza completa de un láser proyectado en la pared o la ola mexicana creada en un estadio) no puede poseer CC y, por tanto, tampoco puede transferirlas. Pero, pese a esto, efectivamente posee identidad a través del tiempo pues superviene siempre en procesos causales o entidades físicas que la deben tener para ser tales (Dowe 2000 105). Por otra parte, según lo que se ha argumentado en este artículo, es fácil apreciar que si uno toma, como base de comparación, un TWG mecánico sublumínico y de baja velocidad, no tendrá identidad en el tiempo pero poseerá CC, las que, no obstante, no podrá transferir. Así, el único rasgo común que subsiste en esta primera comparación es que ambos, PSP y TWG, no transmiten información física en la dirección en que aparecen como movimiento. Ahora bien, en una nueva comparación más atenta de ambos, emergen al menos otros dos rasgos en común: ambos generan compuestos mereológicos con partes espaciales y estas evolucionan como puntos en el tiempo, esto es, ambos son extensiones temporales cuyas partes evolucionan en el continuo espaciotemporal. Pese a estas coincidencias, en la región de lo sublumínico prevalecen las dos diferencias fundamentales que ya se señalaron. Sin embargo, nuestra argumentación en las secciones precedentes muestra que si el mismo TWG mecánico sublumínico se vuelve superlumínico para un dado observador, resultaría ser observado como un todo coexistente y simultáneo para otro observador, y, por lo tanto, como el comienzo de un SWG en este segundo sistema. Si es así, aquella entidad mereológica a la que Dowe le adjudica identidad temporal, un SWG, no la tendría de manera absoluta, pues físicamente podría perfectamente ser descrito como una colección de TWG superlumínicos, es decir, como una entidad con partes espaciales temporalmente desconectadas.

Si analizamos ahora el caso de un PSP superlumínico, por ejemplo, la misma traza de láser pero proyectada imaginariamente sobre la superficie de la luna con cierta velocidad tal que rebase la velocidad c , entonces, ateniéndose a una argumentación análoga a la anterior, no parece haber razón física en la TRE

que impida interpretar el conjunto de puntos de luz de la traza superlumínica como un TWG poseyendo efectivamente CC, al contrario de lo que Dowe piensa. Si esta interpretación se sostiene, entonces somos conducidos a la discusión respecto a la necesidad de adjudicarle identidad en el tiempo a los PSP —rasgo irrenunciable de estos últimos para Dowe— debido a que si efectivamente se puede mostrar que los PSP pueden poseer CC pero no pueden transferirlas, entonces resulta inevitable, a su vez, plantearse si no estamos simplemente ante una manifestación más de los TWG, pues, con excepción de ese rasgo, en todos los demás ambas entidades coincidirían. Dicho al revés, resultaría mucho más complejo defender una diferencia entre dichas entidades basada solo en un supuesto metafísico de carácter primitivo, que aceptar una asimilación que simplifica la ontología de la física sin violar ningún precepto de la TRE. Esta simplificación nos dejaría entonces solo con dos tipos de entidades fundamentales: los PC y las entidades *gerrymandered* como parte de la física¹⁰. Y, desde luego, podríamos prescindir sin ninguna pérdida de la calificación de “basura física” aplicada por Dowe

¹⁰ Es importante señalar aquí que Dowe (1995b) ya había propuesto originalmente una definición amplia de objeto físico como cualquier entidad *gerrymandered* aceptada por la física efectiva, en el contexto de su teoría de CC sin transmisión de marca. Sin embargo, dicha propuesta implicaba un criterio según el cual solo objetos que pueden estar totalmente presentes en un mismo instante pueden contar como objetos o procesos físicos, lo que puede ser objetado desde varios frentes (cf. Kistler 1998, 1999). No obstante, dichas objeciones apuntan a inconsistencias más generales de la ontología de Dowe, en particular respecto a la vaguedad de su definición, la que, según Kistler, impediría hacer una distinción precisa entre los PC y los PSP, y, por ende, afectaría su criterio de posesión de cantidades. Como consecuencia de eso, Kistler propone abandonar el concepto de entidad *gerrymandered* y rescatar la distinción desde la perspectiva del criterio de Salmon, algo diferente al camino que nosotros estamos proponiendo aquí, que, sin rechazar aquel concepto, surge de concentrarnos críticamente en la propiedad de identidad temporal como el criterio crucial de realidad física que organiza toda la ontología de Dowe. Finalmente, es posible imaginar una opción más radical que, sin apelar a la transmisión o posesión de cantidades conservadas, implique un “barrido” de toda pseudo o cuasientidad causal, focalizada solo en el hecho físico que campos relativistas clásicos no permitirían por definición en su evolución procesos que puedan exceder la velocidad de la luz, dadas ciertas condiciones (cf. Earman 2014). Esta opción entonces ya no se haría cargo de dar sentido metafísico a la distinción entre PC y PSP. En tal caso, nos quedaríamos solo con una entidad, esto es, PC, que se ajusta a esas condiciones y nada más. Ya que, como hemos argumentado en otros lugares (cf. Quezada & Pavez 2011, 2022a), un PSP

a una de las últimas entidades y que ahora parece, a lo menos, apresurada y físicamente mal pensada.

REFERENCIAS

- Choi, Sungho. “Causation and Gerrymandered World Lines: A Critique of Salmon”. *Philosophy of Science* 69.1 (2002): 105-117. <<https://doi.org/10.1086/338943>>
- Dowe, Phil. “Causality and Conserved Quantities: A Reply to Salmon”. *Philosophy of Science* 62.2 (1995a): 321-333. <<https://doi.org/10.1086/289859>>
- _____. “What’s Right and What’s Wrong with Transference Theories”. *Erkenntnis* 42.1 (1995b): 363-374. <<https://doi.org/10.1007/BF01129010>>
- _____. *Physical Causation*. Cambridge University Press, 2000. <<https://doi.org/10.1017/CBO9780511570650>>
- Earman, John. “No Superluminal Propagation for Classical Relativistic and Relativistic Quantum Fields”. *Studies in History and Philosophy of Modern Physics* 48 (2014): 102-108. <<http://dx.doi.org/10.1016/j.shpsb.2014.07.005>>
- Hitchcock, Christopher. “Causal Processes and Interactions: What are They and What Are They Good for?”. *Philosophy of Science* 71 (2004): 932-941. <<https://doi.org/10.1086/425236>>
- Kistler, Max. “Reducing Causality to Transmission”. *Erkenntnis* 48 (1998): 1-24. <<https://doi.org/10.1023/A:1005374229251>>

no necesita ser superlumínico para ser tal, esta opción radical nos parece claramente insuficiente para tratar y dar cuenta del problema de la naturaleza de las pseudoentidades en física relativista y en física en general.

- Kistler, Max. *Causalité et lois de la nature*. París: Librairie Philosophique J. Vrin, 1999.
- Quezada, Wilfredo y Pavez, Luis. “Causalidad, pseudocausalidad y medición”. *Revista de Filosofía U. de Chile* 67 (2011): 251-268. <<https://doi.org/10.4067/S0718-43602011000100016>>
- _____. *Velocidad de fase, velocidad de grupo y velocidad de la luz: una guía (causal) para entender sus relaciones*. II Jornada de Fundamentos, Filosofía e Historia de la Física. Universidad de Buenos Aires, Argentina, septiembre 2021. Ponencia.
- _____. *Entidades físicas y basura espaciotemporal. Una reconstrucción crítica de la ontología física de P. Dowe*. XXII Jornadas Rolando Chuaqui en Filosofía y Ciencias. Universidad de Valparaíso, Valparaíso, Chile, octubre 2022a. Ponencia.
- _____. *Ontología física: entidades y simuladores de entidades*. III Jornada de Fundamentos, Filosofía e Historia de la Física. Universidad de Buenos Aires, Buenos Aires, Argentina, octubre 2022b. Ponencia.
- Reichenbach, Hans. *The Philosophy of Space and Time*. Nueva York: Dover, 1958.
- Robb, Alfred. *A Theory of Time and Space*. Cambridge University Press, 1914.
- Salmon, Wesley. “Causality without counterfactuals”. *Philosophy of Science* 61.2 (1994): 297-312. <<https://doi.org/10.1086/289801>>
- _____. “Causality and Explanation: A Reply to Two Critiques”. *Philosophy of Science* 64.3 (1997): 461-477. <<https://doi.org/10.1086/392561>>
- _____. *Causality and Explanation*. Nueva York: Oxford University Press, 1998. <<http://dx.doi.org/10.1093/0195108647.001.0001>>
- Sklar, Lawrence. *Philosophy of Physics*. Oxford University Press, 1992. <<http://dx.doi.org/10.1201/9780429498305>>
- Torretti, Roberto. *Relativity and Geometry*. Oxford: Pergamon Press, 1983.

HOLISMO LOCAL Y CAMBIO SEMÁNTICO EN LA TEORÍA DE KUHN: ARISTÓTELES, NEWTON Y LA INCONMENSURABILIDAD TAXONÓMICA*

LOCAL HOLISM AND SEMANTIC CHANGE IN THE
KUHN'S THEORY: ARISTOTLE, NEWTON AND THE
TAXONOMIC INCONMENSURABILITY

DANIEL LABRADOR-MONTERO
Universidad de Salamanca
Salamanca, España
danilabra@usal.es

<https://orcid.org/0000-0001-5095-1021>



RESUMEN

El objetivo de este artículo es ahondar en la tesis de la inconmensurabilidad taxonómica defendida por Kuhn a partir de los años ochenta. Según el autor estadounidense, la inconmensurabilidad entre teorías, en esta versión más local y moderada, es el resultado del cambio semántico de algunos términos centrales, a los cuales denomina “categorías taxonómicas”. Según Kuhn, dichos términos están interdefinidos holísticamente, de manera que la modificación del significado de uno de ellos implica una redefinición del resto. Para extraer ejemplos de tal holismo local y de la reestructuración semántica que se produce tras una revolución científica, Kuhn acude a la física aristotélica y newtoniana. Sin embargo, aunque los considera ejemplos de gran relevancia y los utiliza como pruebas concluyentes de sus ideas, no los desarrolla en profundidad. Por lo tanto, este

* Este artículo se debe citar: Labrador-Montero, Daniel. “Holismo local y cambio semántico en la teoría de Kuhn: Aristóteles, Newton y la inconmensurabilidad taxonómica”. *Revista Colombiana de Filosofía de la Ciencia* 24.48 (2024): 233-275 <https://doi.org/10.18270/rcfc.4409>

artículo se propone comprobar de manera más detallada si un análisis de tales teorías apoya la tesis de la inconmensurabilidad taxonómica.

Palabras clave: holismo; inconmensurabilidad; categoría taxonómica; cambio semántico; Kuhn; Aristóteles; Newton.

ABSTRACT

This article aims to delve into the concept of taxonomic incommensurability as advocated by Thomas Kuhn from the 1980s onward. According to Kuhn, in this more local and moderate interpretation, the incommensurability between theories results from the semantic alteration of certain central terms, which he refers to as 'taxonomic categories'. He argues that these categories are holistically inter-defined, such that altering the meaning of any one term necessitates a redefinition of the others. To draw examples of such localized holism and the semantic restructuring that ensues following a scientific revolution, Kuhn frequently references the physics of Aristotle and Newton. However, although he identifies these instances as significant, his treatment of them lacks depth. Consequently, this article seeks to thoroughly examine whether an in-depth analysis of these theories substantiates the thesis of taxonomic incommensurability.

Keywords: Holism; incommensurability; taxonomic category; semantic change; Kuhn; Aristotle; Newton.

1. INTRODUCCIÓN

En la compilación *El camino desde la estructura*, que contiene los ensayos más relevantes de las dos últimas décadas de vida de Thomas S. Kuhn (1922-1996), uno de los principales temas es la reflexión sobre la inconmensurabilidad, ya madurada tras las numerosas críticas recibidas. Allí, Kuhn se aleja de los argumentos más idealistas expuestos en *La estructura de las revoluciones científicas* (1970) y obvia, en gran parte, la metáfora del “cambio de visión” para centrarse en la relación entre “estructuras lingüísticas”. La inconmensurabilidad entre teorías aparece, más que nunca, como intraducibilidad (o como la imposibilidad de la traducción total), es decir, como la inviabilidad de encontrar un lenguaje común en el que pueda darse la traducción sin pérdida (Kuukkanen 2012¹ ; Mayoral de Lucas 2017). Por tanto, utilizando palabras de Rorty (1983), la inconmensurabilidad implica que no existe un discurso neutro a través del cual llegar a un acuerdo. Ahora bien, que no haya una medida común no implica que no puedan compararse teorías. La intraducibilidad no cierra el paso a la interpretación hermenéutica y al aprendizaje de otros lenguajes como herramientas necesarias para el historiador (cf. Kuhn 2002a [1982]).

En concreto, según Kuhn (2002a [1982]), en el cambio paradigmático se produce una alteración semántica en una serie de términos que se *interdefinen* —los términos de clase o categorías taxonómicas— que impide cualquier tipo de reproducción de las viejas teorías a través del nuevo vocabulario. Esta inconmensurabilidad local, provocada por el cambio semántico de una serie de términos centrales interconectados, ha sido denominada a menudo “inconmensurabilidad taxonómica”. Este calificativo se debe al concepto de “categoría taxonómica” utilizado por Kuhn. Según este, dichas categorías están en la base no solo de la conceptualización y ordenación de los fenómenos, sino incluso en la elección y determinación de aquellos

¹ Según Kuukkanen (2012), la noción de “cambio de significado” de Kuhn es compleja y plural. Además, argumenta que la ciencia cognitiva respalda la propuesta del cambio de significado kuhniano como aspecto relevante para hacer historia de la ciencia.

objetos y procesos que son considerados o no en una investigación científica. Por tanto, la inconmensurabilidad taxonómica hace referencia a cómo dos científicos educados en dos teorías distintas organizan y clasifican el mundo de manera distinta. Desde esta óptica, un newtoniano no solo entiende y conceptualiza el movimiento de manera distinta a un aristotélico, sino que, incluso, difiere en los fenómenos que cataloga como movimiento.

Teniendo en cuenta esto, el trabajo del historiador consiste en acudir a dicho léxico y aprenderlo, para así poder interpretar las teorías antiguas teniendo en cuenta que la relación holista entre sus términos determina los diferentes significados. Eso mismo hace el propio Kuhn cuando se para a analizar la física aristotélica y la newtoniana, es decir, intentar dar con esas categorías taxonómicas, con esas interrelaciones que hacen que alguien educado en la mecánica newtoniana no pueda establecer una estructura coherente a la obra y argumentación de Aristóteles sin haber hecho un exhaustivo trabajo hermenéutico previo.

En su famoso texto *¿Qué son las revoluciones científicas?*, Kuhn (2002b [1987]) utiliza la teoría de Aristóteles como ejemplo de interdefinición holística de categorías taxonómicas. En algunos pasajes, el autor estadounidense, sin entrar en gran profundidad, señala la interconexión semántica entre los términos esenciales de la teoría aristotélica y los compara y diferencia de nociones semejantes en la teoría de Newton (cf. Kuhn 2002b [1987] 28-31). Lo que quiere demostrar Kuhn es que, fruto de esa interdefinición, el cambio de significado de una de las categorías taxonómicas implica que las demás cambien y que toda la teoría sea modificada o incluso derrocada. Es decir, los términos de clase codependen semántica y teóricamente los unos de los otros. Además, Kuhn no duda en afirmar que este holismo semántico no solo afecta a “cómo se conectan” los términos científicos con la naturaleza, sino incluso que conjunto de objetos y fenómenos se consideran dignos de atención científica. En este sentido, “ejemplos de movimientos que habían sido paradigmáticos para Aristóteles —de bellota a roble, y de la enfermedad a la salud— no eran movimientos en absoluto para Newton” (Kuhn 2002b [1987] 42). Así, para Kuhn, el carácter holista del léxico nuclear de una teoría científica otorga un modo concreto de categorización o conceptualización de la realidad estudiada. De esta manera, una revolución

científica se caracteriza por una redefinición y reconexión de esos términos de clase y “ya que tal redistribución afecta siempre a más de una categoría, y ya que esas categorías se interdefinen, esta clase de alteración es necesariamente holista” (Kuhn 2002b [1987] 43).

Por otra parte, en *Mundos posibles en la historia de la ciencia*, Kuhn (2002c [1989]) realiza un ejercicio parecido con la teoría de Newton, es decir, analiza y afirma la codependencia semántica y teórica de algunas categorías taxonómicas de la teoría newtoniana, como “fuerza” o “masa”. En dicho texto, intenta argumentar que algunos de esos términos de clase no se pueden aprender por separado, sino que, por el contrario, debe darse una introducción integral a tales conceptos junto con el propio aprendizaje de la teoría. De nuevo vuelve acudir a la comparación con la teoría aristotélica, como se puede apreciar en algunas partes (cf. Kuhn 2002c [1989] 88-9) y también con la teoría einsteiniana (cf. Kuhn 2002c [1989] 95). En definitiva, se puede observar que el cambio paradigmático producido entre la física aristotélica y la newtoniana ocupaba para Kuhn un importante papel en este momento de su pensamiento, especialmente en lo relativo al nexo entre inconmensurabilidad y holismo semántico.

De este modo, el estudio del cambio teórico desde el paradigma aristotélico al newtoniano proporciona, según Kuhn, pruebas para aseverar que “con escasas excepciones los términos individuales no tienen en absoluto un significado” (Kuhn 2002c [1989] 99). Es decir, los términos científicos solo adquieren significado en relación con otros y bajo el marco de una teoría. Esto lo afirma mientras rechaza la dicotomía entre términos teóricos y observacionales, por lo que la realidad que experimenta el científico, los fenómenos que selecciona y cómo los categoriza dependen de esas interconexiones semánticas. Respecto a esto Kuhn es contundente: “aquellos que afirman la independencia de la referencia y el significado también afirman que la metafísica es independiente de la epistemología. Ningún punto de vista parecido al mío es compatible con esta separación” (Kuhn 2002c [1989] 99). Por esta razón, el holismo semántico estaría en la base de la inconmensurabilidad, ya que una estructura léxica y un lenguaje científico concretos generan distintos modos de conocer y categorizar los fenómenos y, con ello, otorgan al mundo estructuras metafísicas distintas.

Así las cosas, tomando como eje estas tesis, el presente artículo tiene como objetivo comprobar en qué medida son válidas las ideas de Kuhn acerca del holismo semántico y la inconmensurabilidad entre dichos términos de clase de las teorías aristotélica y newtoniana. Por esta razón, es conveniente destinar un primer apartado a detallar cómo el joven Kuhn, con tan solo 25 años, se enfrentó al estudio de la física aristotélica en un intento de descubrir el progreso científico desde la teoría del griego a la mecánica newtoniana. Sin embargo, tal acercamiento a la obra de Aristóteles provocó que experimentara por primera vez la inconmensurabilidad entre teorías. Lejos de encontrar el desarrollo gradual de una ciencia, Kuhn halló dos mundos distintos: el aristotélico y el newtoniano. Este fue el germen de sus ideas acerca de las revoluciones científicas y la inconmensurabilidad (Wray 2021).

Posteriormente, se analizarán algunos aspectos de la obra de Aristóteles, para después pasar a la de Newton, intentando establecer conexiones más profundas y extensas de las que se presentan en los textos de Kuhn, pues el propio pensador americano reconoce que sus observaciones a este respecto son “simplificadas o incompletas” (Kuhn 2002b [1987] 31). Es necesario tener en cuenta que, aunque con pretensiones algo distintas, recientemente se ha publicado un artículo en el que se evalúa la cuestión de la inconmensurabilidad local entre la física newtoniana y la teoría de la relatividad (Cadavid-Ramírez & Arias-Vélez 2021), algo que ya Kuhn había tratado sucintamente en 1962 en *La estructura* (cf. Kuhn 1970 149). Así, este artículo complementa tal estudio. Finalmente, en un apartado posterior se revisarán, en el marco de la teoría kuhniana, las implicaciones filosóficas del examen realizado en los dos apartados anteriores.

Por último, es oportuno acabar este apartado introductorio haciendo una aclaración. Mucho se ha escrito sobre la transición entre la teoría aristotélica y la mecánica clásica sobre las revoluciones copernicana y newtoniana que conllevaron el destronamiento de la astronomía del filósofo griego y sobre las numerosas aportaciones de grandes pensadores a lo largo de los veintiún siglos que separan a un paradigma del otro. Por ello, con el análisis que aquí se presenta no se quiere contribuir a una “historia de genios” en la que Aristóteles y Newton sean los únicos nombres resaltados en esa larga trayectoria. Desde luego, la llamada revolución newtoniana

ni acabó ni comenzó con Newton. Como explica Kuhn en *La revolución copernicana* (1996), dicha revolución conceptual comenzó con Copérnico, fundamentalmente, y culminó con la asimilación del universo newtoniano. De ahí la importancia de este inciso para evitar lo que Voltaire imputó a la sociedad inglesa del siglo XVIII: tratar a Newton como al “Hércules de la fábula, a quien los ignorantes atribuían todas las hazañas de los otros héroes” (Voltaire 2010 [1734] 63). Por ende, no se quiere subestimar, por ejemplo, la importancia del debate medieval acerca de la obra de Aristóteles —el cual, como dice la famosa frase de Pico della Mirandola, “sería mudo sin Tomás de Aquino”, en clara referencia a la aportación de los comentaristas— ni las imprescindibles aportaciones de Galileo, Kepler, Hooke, Descartes o Huygens, entre otros, pero resulta del todo fundamental centrarse en esos dos autores para orientarse a los objetivos específicos de este trabajo. En cambio, el presente análisis conceptual-terminológico recoge, en cierta medida, la crónica de un cambio en el que la historia se introduce en la semántica. Y es que Aristóteles era un amplio conocedor de las ideas de su mundo, y eso queda plasmado, con la particularidad de una excelencia sin igual, en sus conceptos y en el uso que da a estos. Lo mismo podemos decir de los planteamientos de Newton, que recogen su deuda (a veces de forma reconocida y otras veces no) de los destructores de la física aristotélica y de los impulsores de la revolución moderna de la ciencia.

2. LA EXPERIENCIA ARISTOTÉLICA DE KUHN

La elección de este estudio de caso para evaluar la tesis de la inconmensurabilidad taxonómica no es casual. Kuhn se acercó a la teoría aristotélica en 1947 en un intento de comprender los orígenes históricos de la mecánica newtoniana. Esto supuso un punto de inflexión en su pensamiento pues le llevó a “vislumbrar” por primera vez la inconmensurabilidad (Hoyningen-Huene 2015). Kuhn describe en tres ocasiones lo que algunos autores, como George Reisch (2016), no dudan en llamar “epifanía aristotélica”: la primera de ellas en 1977, en el prefacio de su *The Essential Tension* (cfr. Kuhn 1993a 11-14), la segunda en su ensayo de 1987; *Qué son las revoluciones*

científicas? (Kuhn, 2002b [1987]), y, en última instancia, en la entrevista *Una conversación con Thomas Kuhn*, publicada en *El camino desde la estructura* (Kuhn 2002e).

Lo más importante es que en los tres casos Kuhn habla de su “experiencia aristotélica” como una revelación. Paweł Jarnicki y Hajo Greif (2022) cuestionan la importancia que le da Kuhn a su estudio sobre historia de la mecánica y argumentan que infravalora el impacto que tuvo en su pensamiento la lectura de la teoría de Ludwik Fleck en 1949. En cualquier caso, es evidente que para Kuhn el estudio histórico que realizó fue un elemento determinante para que decidiera cambiar su vida académica y dedicarse a la historia de la ciencia para, finalmente, poder hacer filosofía (Wray 2021 18). Por ello, es oportuno detenerse brevemente en qué obtuvo Kuhn de su acercamiento a la física de Aristóteles y qué cuestiones examina. Así, posteriormente, se podrá mencionar cómo este artículo completa su análisis.

De los textos citados de Kuhn se concluye que, según él, su estudio histórico de Aristóteles le llevó a apreciar por primera vez los siguientes aspectos relevantes para su filosofía de la ciencia: 1) la autonomía semántica y epistémica de las teorías científicas; 2) la interrelación holística entre sus componentes; 3) la inconmensurabilidad; 4) que el desarrollo de la ciencia es discontinuo, no acumulativo y no implica aproximación a una verdad absoluta; 5) que las revoluciones científicas implican cambios súbitos en la organización epistémica de los fenómenos, en la categorización de la realidad y en los objetivos de una ciencia concreta.

Ahora bien, ¿cómo llegó a tales tempranas conclusiones a través de dicho estudio? Kuhn (2002b [1987] 26) confiesa que, en un primer momento, se sintió frustrado y juzgó a Aristóteles como “un físico terriblemente malo”. Sin embargo, esto se debía a que estaba leyendo al estagirita desde las categorías newtonianas. No había aprendido adecuadamente su léxico ni había profundizado en su contexto filosófico y en los verdaderos propósitos de su obra. Estaba siendo testigo directo de los efectos de la incomensurabilidad. Sin embargo, cuando se esforzó y realizó un arduo ejercicio heurístico, todas las piezas se ordenaron repentinamente y se convenció de que esa transformación súbita que él experimentó es, en lo esencial, lo que experimentan los científicos en una revolución (Kuhn 2002b [1987] 27). El punto de inflexión fue comprender algunos aspectos de la tradición aristotélica, especialmente que el

proyecto de Aristóteles era entender el cambio de cualidad en general (Kuhn 1993a 11), pero también necesitó aprender el significado de algunos términos esenciales y la importancia de su cosmología para su mecánica.

Partiendo de esta descripción, Kuhn analiza brevemente solo algunos aspectos que demuestran, a su juicio, la interdefinición de las categorías taxonómicas en la teoría de Aristóteles. En dicha interdependencia radicaría la causa de la inconmensurabilidad pues en una revolución dichos “términos de clase” se ven modificados en conjunto necesariamente. Así, Kuhn evalúa sucintamente 1) la “física cualitativa” de Aristóteles y la diferencia de significado entre su término *movimiento* y el de Galileo o Newton; 2) la imposibilidad de incluir el vacío en la teoría del griego. Analizando estos dos aspectos, intenta mostrar la interconexión holista entre los distintos elementos de la física aristotélica, argumentando que el cambio de significado de uno de ellos implica un “efecto dominó” sobre los otros. Este holismo semántico estaría en la base de la inconmensurabilidad entre la teoría aristotélica y la newtoniana, ya que los términos centrales tienen distintos significados en una teoría y en otra. Por este motivo, es necesario hacer un análisis más integral que el de Kuhn y explorar en qué medida tiene razón y son acertadas sus conclusiones.

3. LA FÍSICA ARISTOTÉLICA: UNA TEORÍA ALREDEDOR DEL CONCEPTO DE MOVIMIENTO

3.1. CAMBIO Y MOVIMIENTO

Según Kuhn, la física-cosmología aristotélica destaca por su carácter holista. De hecho, él explica su experiencia intentando comprender la doctrina del estagirita (Kuhn 2002b [1987] 26-28) y cómo en un primer intento fracasó en su misión precisamente porque en esa inicial tentativa no fue capaz de obtener una visión de conjunto, es decir, no había aprendido las interconexiones semánticas entre los términos centrales de la teoría aristotélica. Sin embargo, continuó esforzándose y una vez logró configurar dicho rompecabezas de interrelaciones la comprensión de

la teoría fue tan abrupta como lo es, según él, la transformación que experimenta el científico que se instala en un paradigma nuevo:

Súbitamente, los fragmentos en mi cabeza se ordenaron por sí mismos de un modo nuevo, encajando todos a la vez. Se me abrió la boca porque de pronto Aristóteles me pareció un físico realmente bueno, aunque de un tipo que yo nunca hubiera creído posible. Ahora podía comprender por qué había dicho lo que había dicho [...]. Este tipo de experiencia —las piezas ordenándose súbitamente por sí mismas y apareciendo juntas de un modo nuevo— es la primera característica general del cambio revolucionario [...]. (Kuhn 2002a [1982] 27)

Según Kuhn, el eje de la teoría aristotélica es el término movimiento, al que de forma más directa o indirecta refieren los demás conceptos de especial relevancia. Es menester, por tanto, comenzar con un análisis de este para después reflexionar sobre el resto de los conceptos que conformarían, en términos de Kuhn, una taxonomía léxica. Como se ha señalado, el físico estadounidense realiza un breve examen a este respecto en *¿Qué son las revoluciones científicas?* (2002b [1987] a modo de ejemplificación de lo que supone conceptual y lingüísticamente una revolución científica. En este apartado se intentarán dar unos pasos más de los recorridos en dicho texto.

En primer lugar, como el propio Kuhn destaca (2002b [1987] 28), el concepto de movimiento es lo primero que resulta problemático a la hora de introducirse en la física aristotélica. Mientras que en las mecánicas galileana y newtoniana el movimiento se restringe al “cambio de posición”, Aristóteles partía de un concepto mucho más amplio de movimiento, en ocasiones sinónimo de “cambio en general”². El movimiento es un tipo de cambio que engloba tanto el movimiento local como

² Entendamos por este el cambio de un “sujeto” a otro “sujeto”, ya que la corrupción y la generación quedan excluidas como movimientos pues “no hay movimiento fuera de las cosas” (*Física* III I 200b32) y un movimiento necesita un “desde dónde” y un “hacia dónde” (cf. *Física* V I) y el no-ser no implica un dónde.

el cambio cuantitativo y cualitativo. Por lo tanto, el significado de movimiento está irremediabilmente ligado al de los términos *lugar*, *cualidad* y *cantidad*. En este artículo, el interés lo demanda el concepto de lugar. Sin embargo, antes resulta necesario hacer otro tipo de aclaraciones.

Para Aristóteles el objeto de la física o filosofía natural son los *seres mutables*, que están sometidos a cambio y son independientes. Esto otorga al “movimiento” un papel central en su comprensión de la naturaleza. Para el filósofo griego, la realidad del movimiento es algo que no se puede dudar, algo que obtenemos de forma evidente por *epagogé* (experiencia o inducción) (*Física* I II 185a13-14). Tanto es así que incluso la naturaleza (*physis*) es definida en relación con el movimiento³: “la naturaleza es un principio y causa del movimiento o del reposo en la cosa a la que pertenece primariamente y por sí misma, no por accidente” (*Física* II I 192b21-22). De esta forma, lo natural es aquello que guarda en sí un “principio de movimiento” —y, por ello, también de reposo—, mientras que los artefactos (los productos del arte) no tienen dicha “tendencia natural al cambio”, es decir, no tienen la capacidad motriz *en sí*, sino que se mueven accidentalmente.

Como se señalaba, el movimiento es un extenso subconjunto del conjunto que engloba a todos los cambios, y si hay algo novedoso en este concepto de cambio con respecto a sus antecesores es el remplazo de una simplista estructura diádica del cambio entre dos contrarios por una triádica. Con esto, el filósofo de Estagira quiere remarcar que *en todo cambio algo se mantiene*, contradiciendo el “todo fluye, nada permanece” de Heráclito, y eso que persiste es el “sujeto” o “substrato” (*hypokeímenon*) (cfr. *Física* I VII). De esta forma, Aristóteles quiere mostrar que el cambio no es, o al menos no solo es, el paso de un ser a su contrario, esto es, a su no ser (*aniquilación* o *destrucción*), sino que puede darse un “cambio de estado” de una misma cosa. Esto es el movimiento a nivel fundamental, un cambio de estado que se da de forma procesual en un sujeto o substrato que se mantiene. Como señala Alexandre Koyré

³ Aristóteles indica que si ignorásemos lo que es el movimiento ignoraríamos lo que es la naturaleza (*Física* III I 200b14-15).

(1980 11) en sus *Estudios galileanos*, para Aristóteles el movimiento “es un proceso, un devenir en el cual, y mediante el cual se constituyen, se actualizan, se realizan los seres”. Se puede apreciar, por tanto, lo íntimo de los términos *cambio* y *movimiento*.

Como explica Kuhn, el principal rasgo de los términos de clase es que se interdefinen unos a otros. Así, de la misma manera que *physis* se define en función de *movimiento*, este último lo hace en relación con *naturaleza*, es decir, existen los movimientos que son *por naturaleza* y los que no lo son (*violentos*). Esto, a su vez, implica mantener un supuesto muy importante, a saber, que *todo lo que se mueve es por algo*, sea “por sí mismo” o por otro, reservándose lo primero para lo natural⁴. En definitiva, que una cosa tenga el principio de movimiento de forma esencial —que se mueve por sí misma— es lo que la hace natural, mientras que el movimiento es “por naturaleza” cuando se rige por tal principio. También el reposo se define a través de la noción de naturaleza, pues algo está en reposo, en sentido estricto, cuando ha alcanzado su lugar natural. Ahora bien, solo puede estar en reposo aquello que tiene el principio de movimiento, luego el reposo es la “privación del movimiento en aquello que puede recibir movimiento” (*Física* v II 226b15-16), pero en la medida en la que el fin del movimiento es *acto*. Esta sutil diferencia nos muestra que el reposo no es simple inmovilidad como en la mecánica clásica del siglo xvii.

Ahora sí se puede entender la definición aristotélica de movimiento (*kinesis*) como “la actualidad de lo potencial, cuando al estar actualizándose opera no en cuanto a lo que es en sí mismo, sino en tanto que es movable” (*Física* III I 201a28-29). Aristóteles quiere resaltar el carácter procesual del movimiento y, por ello, utiliza la palabra *energeia* (actualización) y no *entelecheia* (actualidad) (cf. Chen 2009). Como señala Pierre Aubenque (1981), en la definición general de *entelecheia* hay referencia al movimiento, siendo esta el acto opuesto al movimiento en cuanto actividad, pues sería más bien el resultado del mismo, esto es, “el modo de ser de lo inmóvil, pero de un inmóvil *que ha llegado* a ser lo que es” (422). De esta forma, el movimiento natural es una actualización hacia un *telos* al que un objeto natural

⁴ Lo artificial también se puede mover por sí mismo, pero no esencialmente sino solo por accidente, esto es, de forma circunstancial.

tiende por naturaleza, o sea, la causa de su movimiento es su fin, que no es otro que lograr su *eidós*. Como se puede apreciar, la causa eficiente, formal y final coinciden en este tipo de movimiento (*Física* II VII 198a25).

Esta tendencia natural a un fin explica lo que Kuhn adecuadamente denomina asimetría en el cambio (2002b [1987] 30), algo que resulta evidente en los cambios cualitativos pues una bellota se desarrolla naturalmente en un roble, pero no al revés. Esto explica, que, por coherencia del sistema, el movimiento local (*phorá*), como un tipo de cambio, tenga también que ser asimétrico. Por tanto, una piedra, por ejemplo, se mueve naturalmente hacia el centro del universo, pero no sucede lo contrario, pues tal movimiento al centro se trata de un tránsito tan natural como que de una semilla surja una planta.

3.2. LUGAR, ORDEN Y FINITUD

De lo anterior se deduce que para esta teoría teleológica no solo se necesita una causa o principio natural del movimiento en lo movido y movable, sino que además se requiere un *lugar natural* para cada cuerpo que se mueve por naturaleza, es decir, para cada ser que pertenece a la *physis*. Si tenemos en cuenta la influencia de los filósofos antecesores de Aristóteles, así como esta necesidad de describir un lugar natural para cada ser, era de esperar que los elementos (*stoicheia*) que componen el cosmos sean reducidos. De esta forma, Aristóteles los reduce a cinco, distribuidos entre dos esferas: la supralunar y la sublunar. La primera, incorruptible, inmutable y perfecta, queda reservada para el éter; la segunda, corruptible, mutable e imperfecta, es compuesta por los cuatro elementos que ya había introducido Empédocles, a saber, tierra, agua, aire y fuego por orden de cercanía al centro de la Tierra, que a su vez coincide con el centro del universo⁵ (cf. *Acerca del cielo* 296b). Como es lógico,

⁵ Esta es la forma que tiene Aristóteles de establecer un marco absoluto desde donde definir el movimiento ascendente (movimiento que aleja al cuerpo del centro del universo) y descendente (movimiento hacia el centro).

puesto que el movimiento es la actualización de un ser hacia su fin, a cada elemento en su simpleza le corresponde un solo movimiento natural y simple⁶, por ejemplo, al aire un movimiento rectilíneo ascendente hasta situarse debajo del área ocupada por el fuego, mientras que a la tierra le corresponde un movimiento rectilíneo descendente hacia el centro del universo, donde encontraría su lugar de reposo. En cambio, los cielos (la esfera supralunar), al estar compuestos por éter y al ser inmutables, sufren paradójicamente algo que se puede denominar *movimiento sin cambio*: el movimiento circular, que ni tiene principio ni fin ni tampoco un contrario al que oponerse. Lo realmente interesante es ver cómo en esta teoría física el movimiento rectilíneo está ligado al cambio, a la imperfección, pero a la vez a la búsqueda de la perfección, del telos, del reposo. El estado inicial es una “privación” (*stéresis*) del estado final, siendo este último el *eidos*. En cambio, en el movimiento circular no se puede apreciar un comienzo o un fin al que llegar, es ya de por sí perfecto. En las siguientes líneas se argumentará que la separación en dos esferas no es algo aleatorio, al igual que tampoco lo es la eternidad de este movimiento circular de los cielos. Sin embargo, dicha ruptura no solo se debe a cuestiones de coherencia de la teoría, como bien señala Kuhn (1996) en la *Revolución copernicana*, sino que Aristóteles, como gran conocedor de las teorías de su época, toma la idea extendida de la “majestad de los cielos” en comparación con la fragilidad e imperfección terrestre (cf. *Acerca del cielo* 270b).

Los conceptos de movimiento natural y su indisociable o incluso inherente noción de lugar natural acarrear muchas consecuencias. Para empezar, tales ideas no podrían pensarse en una concepción infinita del universo. El hecho de que el movimiento de los cuerpos naturales sea unívoco y orientado a un lugar natural específico no se puede conciliar con un cosmos infinito. El movimiento teleológico necesita de un espacio limitado, pues ¿dónde estaría, si no, el centro del universo al

⁶ Si algo fuera afectado por más de un movimiento (natural y forzado), entonces estos se obstruirían o estorbarían. Esto queda muy lejos de la mecánica newtoniana donde un movimiento puede ser causa de la composición de dos o más fuerzas.

que tienden los cuerpos pesados? En un mundo infinito no habría centro. Por otro lado, “la noción de «lugar natural» traduce una concepción del orden puramente estática” (Koyré 1980 9), es decir, la tendencia al movimiento de los cuerpos naturales se debe precisamente a la búsqueda de su fin, a ordenarse en sus respectivos lugares, lo cual equivale a alcanzar el reposo. Si todo estuviera en orden, todo descansaría, solo habría lugar para el movimiento forzado. En consecuencia, el movimiento conlleva la asunción del desorden. Se puede apreciar con facilidad entonces que, si todo movimiento natural es un retorno al orden, hacia aquel lugar violado, el movimiento no puede ser un hecho que se extienda hacia el infinito (lo cual tendrá consecuencias para la noción de tiempo): “*contra naturam potest esse perpetuum*” (Koyré 1980 10). Ahora bien, si tomamos en conjunto el sistema cósmico, todo movimiento tiene como causa otro movimiento, lo que haría de este fenómeno un hecho eterno, y de ahí que Aristóteles tenga que recurrir a una esfera supralunar separada donde el movimiento sea circular, uniforme y eterno —pues el movimiento circular es el único que puede “perdurar indefinidamente” en una extensión espacial limitada— como causa del resto de los movimientos de la esfera sublunar (cfr. Koyré 1980 10-14). De esta forma, por contacto, ya que no hay causa de movimiento a distancia (cf. *Física* 202a y 243a32-245b1), la esfera celeste más próxima a la esfera sublunar trasmite el movimiento a esta⁷.

Una vez visto lo anterior, resulta fácil hacerse una idea de la relevancia del papel que ocupa el concepto de lugar en la física aristotélica. Si hay algo que evidencia la definición de movimiento natural es que el lugar “ejerce un cierto tipo de poder” (*Física* IV I 208b10). El lugar, lejos de ser una extensión inerte, neutra, homogénea e inactiva, como el espacio de la mecánica newtoniana, desempeña una función

⁷ En Aristóteles encontramos un *principio de acción y reacción* particular y aplicado a su teoría teleológica y a sus conceptos: “[...] actuar sobre lo movable en cuanto tal es precisamente moverlo; pero lo moviente hace esto por contacto, de tal manera que al mismo tiempo experimenta también una modificación” (*Física* III II 202a5-7). El cielo, como motor o moviente que trasmite el eidos en este caso, por contacto comunica movimiento al mundo sublunar, pero no se ve influenciado por lo movido, es decir, no habría reacción alguna.

esencial. El movimiento natural es tal, precisamente, porque hay un lugar hacia el cual tienden los cuerpos naturales, y el concepto de lugar tiene sentido porque existe el movimiento en el cosmos. La suposición de que existen lugares naturales es una *conditio sine qua non* para desrelativizar el movimiento, es decir, en función del lugar natural, y en particular tomando el centro del universo como principal referencia, se puede afirmar que un cuerpo se mueve hacia arriba (alejamiento del centro terrestre) o hacia abajo (acercamiento al centro terrestre). Lugar y movimiento poseen definiciones indisociables. Pero hay algo más que resulta indispensable respecto a las características que ha de tener el lugar. El lugar es “el límite del cuerpo continente ‘que está en contacto con el cuerpo contenido’” (*Física* IV IV 212a5-6), y es inmóvil y separable del contenido (pues sino se movería con el cuerpo que contiene). Pero lo llamativo de esta definición es que se considera al lugar un cuerpo y esto implica la negación definitiva del vacío. Como consecuencia cosmológica obtenemos un universo *plenum*. Ahora bien, ¿por qué el *vacuum* resulta discordante con el resto de la teoría? La respuesta la podemos dividir en cuatro argumentos principales: 1) respecto al movimiento forzado y transmitido: “ninguna cosa singular podría moverse si existiera el vacío” (*Física* IV VII 214b29-30), porque el movimiento, por ser un proceso, necesita como causa el contacto del motor o moviente, y se rompería la cadena entre las diferentes esferas cósmicas si existiera el vacío. 2) El movimiento natural es dirigido y en el vacío es indiferenciable el *hacia dónde*, el *telos*. 3) El movimiento natural se produce por la vía más corta, tan rápido como el medio lo permite, pero en el vacío se daría una velocidad infinita y, por tanto, movimiento instantáneo (Koyré 1980 13). 4) Por otro lado, la posición —que no deja de ser una categoría cualitativa más de las sustancias— no puede entenderse separada de la materia, o sea, si la posición es algo inherente a los cuerpos no puede haber espacio o extensión sin cuerpo.

La concepción de un universo lleno es coherente, a su vez, con la necesidad de la finitud del universo, pues en un cosmos infinito también la materia sería infinita, algo impensable para el pensador de Estagira. Además, si el espacio fuera vacío, y se permitiera el movimiento a través de este, dice Aristóteles (*Física* IV VIII 215a19-21), sería necesario un movimiento rectilíneo hacia el infinito al no tener lugar y razón

por la que reposar el cuerpo. Como señala Kuhn (1996 131), el vacío y la infinitud solo serán elementos posibles una vez se le dé movimiento a la Tierra tras Copérnico.

3.3. EL TIEMPO ARISTOTÉLICO

Según Jesús Conill (1981), tiempo y movimiento son conceptos totalmente indisolubles en la física aristotélica y no se pueden entender por separado. Aristóteles comienza su reflexión sobre el tiempo (contenida fundamentalmente en *Física* iv 10-14) aseverando que, si bien el tiempo no puede identificarse con el movimiento, “no hay tiempo sin movimiento” (219a1). El tiempo como tal es siempre igual: mientras que el cambio se da más rápido o más despacio en referencia al tiempo, el tiempo no es rápido ni lento, en él no cambia nada (cfr. *Física* iv x). Ahora bien, el tiempo es inherente al movimiento hasta tal punto que experimentar movimiento implica experimentar el tiempo y, a la inversa, “la percepción del tiempo incluye siempre la del movimiento” de forma simultánea (Conill 1981 122). Parece que el estagirita incluye un carácter psicológico al tiempo propio de la percepción humana del movimiento, lo cual no elimina la realidad del cambio. Como indica Van Fraassen (1970), no podemos percibir el tiempo en sí, lo que lo hace conceptualmente dependiente del movimiento, que sería la condición necesaria y suficiente para la percepción del primero. Esto no implica que uno sea causa del otro, sino que simplemente se dan en correlación.

La conclusión a la que llega Aristóteles es que el tiempo es “algo perteneciente al movimiento” (*física* 219a10), pues el tiempo no se puede identificar sin el movimiento, pero tampoco es una sustancia, es decir, algo independiente, separable. “El tiempo solo se da en conexión con el ser que es en el tiempo”, es “predicado”, nunca “sujeto” (Conill 1981 132), pero esto no lo convierte en un mero accidente, ya que los accidentes no son algo necesario, es decir, se puede dar la sustancia sin ellos, y el tiempo es imprescindible para el movimiento. El movimiento, se señalaba, siempre

es un cambio de trayecto continuo⁸, es decir, la “magnitud”⁹ es continua y, como consecuencia inmediata, el movimiento y el tiempo también¹⁰. Esta es la razón por la que un movimiento se puede considerar uno y único, precisamente en función del tiempo. Ahora se puede apreciar con claridad que el significado del concepto de tiempo es indisoluble del de movimiento como cambio procesual en el que un sujeto o sustrato se mantiene.

Como conclusión necesaria de todo lo anterior se extrae que, si el tiempo es indisoluble del movimiento y si, como se vio, hay una concepción estática del orden y, por lo tanto, una concepción pasajera del movimiento inherente a su teleología donde el reposo es el *telos* de todo movimiento, entonces el tiempo ha de ser obligadamente finito, pues en un mundo ordenado no habría tiempo. Esto, sin embargo, conlleva una nueva dificultad: si el tiempo acompaña siempre al movimiento, entonces el movimiento circular de los cielos que se considera eterno supondría un tiempo ilimitado. He aquí otra de las razones por las que Aristóteles necesita recurrir a *la majestad de los cielos*, pues la perfección está en los cielos imperturbables, y ser en el tiempo conlleva estar a merced de sus efectos (de cambio, corrupción, etc.). En conclusión, *lo eterno no es en el tiempo*.

4. LA FÍSICA NEWTONIANA: UNA REVOLUCIÓN CONCEPTUAL

Se ha podido ver cómo en la física aristotélica diferentes conceptos centrales se inter-definen y, por ende, son dependientes unos de otros debido a sus relaciones teóricas.

⁸ Se puede definir “continuo” en la teoría aristotélica como la propiedad de que los límites o extremos de dos cosas se identifican formando una unidad.

⁹ A veces referida como lugar y otras, equivocadamente, como espacio, pues en Aristóteles no hay una teoría del espacio como tal, sino una teoría de la posición o del lugar (Ross 1957).

¹⁰ La argumentación es más bien al revés, puesto que el movimiento tiene que ser entendido como continuo necesariamente, entonces el lugar (que es cuerpo) ha de serlo también. En un mundo ordenado, en reposo, en acto, el movimiento no sería necesario y, por tanto, tampoco la continuidad.

Según Kuhn, sucede lo mismo en la mecánica newtoniana, solo que esas relaciones teóricas en ocasiones son fruto de que los conceptos están insertos en leyes. A continuación, se analizarán algunas interrelaciones clave en la física del científico británico. Para ver cómo tras la revolución newtoniana se produce un cambio semántico tal que no es posible utilizar los mismos términos de forma indiferente, será necesario que se tenga en cuenta lo expuesto sobre la física aristotélica. Por ejemplo, mientras que para Aristóteles el movimiento es el concepto absoluto, en el sentido de que a través de él se definen el lugar y el tiempo, en Newton tenemos todo lo contrario, un espacio y un tiempo absolutos que enmarcan la noción de movimiento. Esta puede ser una de las razones que llevaron a Newton a comenzar sus *Principia* con ciertas definiciones necesarias para poder entender las leyes del movimiento. En el siguiente apartado se profundizará en los conceptos de espacio, tiempo y movimiento. Posteriormente, en la siguiente sección, se evaluará la relación que remarca el propio Kuhn entre los conceptos de masa y fuerza.

4.1. UN ESPACIO Y TIEMPO INDEPENDIENTES PARA UN MOVIMIENTO VERDADERO

El espacio es independiente de los cuerpos y el tiempo fluye uniformemente sin importar si algo sucede en el mundo. Solo a través de estas dos premisas se puede entender la naturaleza del movimiento verdadero para Newton¹¹, como señala Robert

¹¹ Sin embargo, hay autores como Belkind (2013) o Van Dyck y Verelst (2013) que señalan que tales premisas tienen un enfoque fundamentalmente metodológico en el *Scholium* de los *Principia*, mientras que aquellos que se centran en *De Gravitatione* destacan el aspecto teológico y metafísico de estas. Con todo, Janiak (2013) asegura que incluso en *De gravitatione* la naturaleza de tales argumentos es también metodológica. Sin embargo, la postura de Janiak no daría cuenta de la cantidad de argumentos y afirmaciones metafísicas que hay en tal escrito, como la aseveración de Newton de que el espacio no solo existe, sino que es la precondition para la existencia de cualquier otra cosa (cf. Newton 2004 25). En dicho texto, Newton arguye que el espacio es eterno al igual que Dios, pues si alguna vez Dios hubiera existido sin espacio hubiera existido en ninguna parte, lo cual es imposible.

Rynasiewicz (2014). Esta sustantivación del espacio y del tiempo suponía arremeter contra visiones muy arraigadas desde la Antigüedad y que no habían desaparecido en el siglo xvii. El *plenum* mantenido durante siglos y reiterado por el influyente sistema cartesiano expuesto en *Le monde* (Descartes 1986 [1664]) era puesto en cuestión por un espacio no cualitativo e incorpóreo, vacío. De hecho, según Friedman (2012 342), para la construcción de la metafísica¹² en la que Newton insertaría su teoría física, “lo más importante fue rechazar decisivamente la identificación de Descartes de la materia con la extensión y defender, en consecuencia, el concepto de espacio absoluto (vacío) que existe con anterioridad e independientemente de la materia”, algo que se explicita contundentemente en su no publicado *De gravitatione* (Newton 2004 22). Solo teniendo en cuenta esto se puede entender el comienzo de la *Decimocuarta carta* de Voltaire titulada *Sobre Descartes y Newton*: “un francés que llega a Londres encuentra las cosas muy cambiadas en filosofía, como en todo lo demás. Ha dejado el mundo lleno; se lo encuentra vacío” (2010 [1734] 60).

Dos aspectos fundamentales tuvieron que cambiar, señala Koyré (1965) en sus *Études newtoniennes*, para que se produjera una revolución de esta índole: la destrucción del concepto de cosmos y la geometrización del espacio. Con “destrucción del cosmos”, que puede parecer una expresión algo confusa, Koyré se refiere al abandono de conceptos como perfección o armonía y, lo más importante, al rechazo de las causas formales y finales, además de un cambio de enfoque de lo que era un mundo en devenir y cuya esencia es el cambio por el “mundo del ser” (Koyré 1965 6-8). En cuanto a la segunda, Koyré se refiere al establecimiento de una visión realista de la geometría euclidiana. Como se puede apreciar, el cambio conceptual es de gran magnitud y, por ello, según la argumentación kuhnianna, el término *movimiento* muta de significado junto con el resto de los que formarían el léxico o esquema conceptual, como *espacio*, *tiempo*, *fuerza* o *masa*. Los dos últimos serán tratados de-

¹² Los estudios acerca de las implicaciones filosóficas, especialmente metafísicas, de la teoría de Newton se han multiplicado en las últimas décadas. Algunas publicaciones relevantes son las Janiak (2008, 2010, 2013), Stein (2002), Belkind (2021) y Schliesser (2012, 2013).

talladamente en el siguiente apartado pues es menester mostrar con anterioridad la ligazón teórica y semántica entre el movimiento, el espacio y el tiempo newtonianos.

Un elemento esencial para entender el holismo local que destaca Kuhn en la teoría de Newton y la necesidad de introducir estos conceptos adjetivados de absolutos —tan criticados, a menudo, por sus implicaciones filosóficas— es que el físico británico nunca quiso dar por hecho ninguno de esos significados, sino que él los definía dentro de un marco concreto de leyes físicas. Si se obvia ese escenario teórico, no tiene sentido preguntarse por la validez de dichas nociones. “Las preguntas correctas son: ¿cuáles fueron las definiciones de Newton de «espacio absoluto», «tiempo absoluto» y «movimiento absoluto»? Y, ¿cómo funcionan esas definiciones en su teoría física?” (DiSalle 2004 36) (traducción propia).

Belkind (2013 495) afirma que “los argumentos de Newton en el *Scholium* se dirigen principalmente contra la definición de Descartes del movimiento verdadero”. A diferencia de Descartes, para Newton el movimiento verdadero es el absoluto, en contraposición con el relativo, y ha de darse de forma obligada a través del espacio absoluto. Aunque algunos han querido matizar la “absolutez” de dicho espacio (cf. Musso 2014), lo cierto es que el propio físico británico lo señala con rotundidad: “el espacio absoluto, por su propia naturaleza y sin relación con cualquier cosa externa, permanece siempre igual e inmóvil” (Newton 1974 [1687] 6) (traducción propia)¹³. Esta es la definición que aparece en el *Scholium* justo antes de la exposición de las leyes del movimiento, la cual otorga mucha información al lector. En primer lugar, explica que en la naturaleza del espacio absoluto está el ser independiente o separado de cualquier cuerpo. En segundo lugar, es fijo, neutro e inmutable. En contraposición con esto, el espacio relativo es aquel cuya medida se dice en referencia a algún cuerpo o sistema corporal y, por lo tanto, no es independiente de él. El “lugar absoluto” quedaría reservado, en cambio, para aquella porción del espacio absoluto ocupada por un cuerpo, siendo lugar y cuerpo separables como en Aristóteles, pero en este

¹³ Texto original: “Absolute space, in its own nature, without relation to anything external, remains always similar and immovable”.

caso el lugar no es un continente material. Esta fue una de las grandes discusiones filosóficas del siglo XVII y autores como Leibniz, Berkeley o Huygens criticarían la concepción newtoniana duramente por diferentes razones.

La propuesta de Newton acerca del espacio no era aleatoria, algo que señaló acertadamente Albert Einstein (1970 15). Y es que no puede comprenderse sin tener en cuenta que su pretensión era justificar un determinado concepto de movimiento, a saber, uno independiente de todo sistema de referencia relativo y corporal. Su intención era rechazar que la relatividad del movimiento que se describe en la obra de Descartes sea la única posibilidad, para a su vez salvar el concepto de velocidad, el cual, para el británico, no tendría sentido en el movimiento cartesiano (Sellés & Solís 1994 137). El espacio absoluto, como señala Paolo Casini (1971 54), “se trata de una necesidad lógica y ontológica, más que física, sin la cual ningún movimiento sería concebible” en el sistema newtoniano y, además, es prueba de una concepción realista de la geometría (Jammer 1970 129). Ahora bien, esto implicaba, como buen atomista que era Newton, rechazar la identidad entre extensión y materia, y aceptar el vacío tan defendido por los atomistas griegos y tan tildado de contradictorio desde Platón y Aristóteles. De esta forma, no solo rechazaba algunas ideas de sus coetáneos¹⁴, sino que se enfrentaba a los grandes clásicos. Esto no quiere decir que no tuviera claros precursores. Sin ir más lejos, su maestro Isaac Barrow¹⁵ fue un ferviente crítico de la identificación entre materia y extensión que promulgaba Descartes y gran defensor de un espacio ilimitado, fijo, isotrópico y penetrable.

¹⁴ Esto no supone que aquellos que tenían nociones diferentes de movimiento y espacio simplemente obstaculizaran la teoría de Newton. Lejos de eso, los *Principia* no habrían sido posibles sin las aportaciones de Descartes o Huygens, por ejemplo, ni el desarrollo de la física clásica habría sido tan fructífero sin la contribución de Leibniz a las matemáticas. Además, como explica DiSalle (2020), una de las grandes preocupaciones de Newton fue el debate sobre la relatividad del movimiento y profundizó en el tema más que ninguno de sus coetáneos.

¹⁵ Algunos autores (Hall 1992; Strong 1970) defienden que Barrow tuvo una influencia decisiva en la noción euclidiana y realista de espacio absoluto de Newton, mientras que otros dudan de la relevancia de tal impronta (McGuire & Tanmy 1983).

Además de lo expuesto, hay otra razón teórica que requiere de este nuevo concepto de espacio: la ley y el concepto de inercia. Como se formula en la primera ley del movimiento, *un cuerpo tiende a mantener su estado de reposo o movimiento rectilíneo uniforme a no ser que una fuerza exterior lo impida*. Pues bien, ¿acaso podría promulgarse esta ley si no fuera bajo la concepción de un espacio (y también de un tiempo) absoluto e infinito? Desde luego, el concepto de movimiento rectilíneo uniforme y de reposo en sentido absoluto necesita de un sistema de referencia fijo y no relativo (Friedman 1983; Jammer 1970 134; Newton 2004 19; Stein 2002), es decir, independiente del lugar relativo¹⁶. Nada habría en reposo en la Tierra si esta se mueve, aunque un objeto se mantenga en el mismo lugar (relativo) respecto de ella. Pero, además, mantener la afirmación de que un cuerpo en movimiento no perturbado nunca por una fuerza seguirá moviéndose hasta el infinito necesitará, obviamente, de un espacio infinito en el que moverse perpetuamente. Ahora bien, se pueden extraer otras consecuencias de la formulación de la primera ley. En primer lugar, el movimiento es un estado de los cuerpos y no un proceso. Aristóteles tenía razón al considerar que un proceso no persevera si no es sustentado por una cadena causal continua. Si el movimiento no cesa cuando la causa de este ya no lo acompaña es precisamente porque no es un proceso, pues todo proceso, como todo cambio, necesita una causa. El movimiento puede conservarse y persistir porque es un estado, como el reposo (del cual no parece ningún problema su persistencia). En segundo lugar, el concepto de materia adquiere un nivel de importancia que no lo tenía en la física aristotélica, ya que adquiere un principio activo, una “fuerza ínsita”, en los términos de Newton, y, además, deja de ser ese sustrato neutro y gobernado por la forma. Es patente que el concepto de materia tiene que adquirir este estatus y este significado nuevo para poder encajar en la teoría (esto no quiere decir que lo adquiera solo por ello).

¹⁶ Acerca de las dificultades que surgen para una distinción verdadera entre movimiento y reposo absolutos dentro de la teoría newtoniana, véase el ensayo de DiSalle (2004 40-41).

Existe otro gran rasgo diferenciador entre la noción de espacio, pero sobre todo de tiempo, con respecto a las concepciones que se convirtieron en comunes tras Aristóteles. El espacio y el tiempo absolutos no son perceptibles. Es oportuno recordar que el concepto de tiempo tenía en Aristóteles un gran componente perceptual y de acción (de medición). En cambio, en el sistema newtoniano, lo perceptible con los sentidos son el movimiento y el tiempo relativos, mientras que el tiempo absoluto fluye independiente a las cosas. Como explican Achtner, Kunz y Walter (2002 113), el tiempo absoluto es lo que hace objetivas a las leyes y a la ciencia en general. Y es que el tiempo absoluto, señala Li Qu (2014 437), es una “medida trascendente”, fruto de “la imaginación racional y no de la observación empírica”. Esto lo aleja del aristotelismo y lo acerca a Platón:

The relationship between absolute time and relative time is akin to that between the Platonic idea and its corresponding physical object. For Plato, the object is brought into being by ‘imitating’ the idea but cannot achieve the latter’s perfection; for Newton, thereby, relative time imitates absolute time but can never attain it perfectly. (Qu 2014 438)

Así las cosas, para Newton, incluso si no hubiera mundo, si nada fuera en el tiempo, todavía habría tiempo y duración. Pero ¿duración de qué? La respuesta no explícita es “la duración de Dios” (Koyré 1965 103-104). La materia necesariamente ha de estar en el espacio y en el tiempo absolutos, pero estos últimos no necesitan de la materia. En la teoría aristotélica no solo sería inconcebible el espacio sin materia, sino también el tiempo sin ella y, por tanto, sin cambio y movimiento. El tiempo deja de ser algo relativo al movimiento para afirmarse la dependencia contraria.

El tiempo absoluto, aquel que “fluye uniformemente sin relación con nada externo” (Newton 1974 [1687] 6), además de ser independiente, es *absolutamente simultáneo*, es decir, cada momento temporal es en todas partes. De esta forma, es un hecho objetivo y no relativo que dos eventos sucedan al mismo tiempo (Disalle 2004 39). La característica de la independencia y la de simultaneidad son esenciales para comprender el tiempo dentro de la física de Newton. Sin embargo, muchas

críticas filosóficas intentan ir más allá y suponen, por ejemplo, que Newton le da una naturaleza sustancial al tiempo. Sin embargo, lo que hay que evaluar es si esta noción cumple una función necesaria dentro de los *Principia*, lo cual es indudable: el que dos intervalos de tiempo sean iguales, incluso sin que haya nada que los mida, es un requisito para distinguir los movimientos inerciales que cruzan distancias iguales en tiempos iguales de los movimientos acelerados causados por una fuerza impresa (DiSalle 2004 39). El concepto de aceleración necesita, por tanto, de un tiempo constantemente uniforme.

Otra razón por la que el espacio, el tiempo y el movimiento han de ser absolutos para Newton responde a la *regla II* que establece en su *Sistema del mundo*, donde afirma —tras introducir la navaja de Ockham en el seno de la física con la *regla I*— que “deben asignarse las mismas causas, en la medida posible, a los efectos naturales de la misma clase” (cfr. Newton 1974 [1674] 398). Esto implica suponer “la uniformidad de la naturaleza y la validez general de las leyes naturales” (Rossi 1998 219), para lo cual se tiene que evitar el relativismo, y afirmar un espacio y un tiempo absolutos donde fuerzas universales como la inercia y la gravedad puedan ser causas de los mismos efectos, de tal forma que se puedan establecer leyes generales que valgan para todo lugar y momento.

Ahora resulta manifiesto que el significado de *movimiento absoluto* como el *traslado de un cuerpo de un lugar absoluto a otro* es dependiente de las nociones de espacio absoluto y de tiempo absoluto. Con Newton, el movimiento y el reposo ya no existen en los cuerpos, ya no son cualidades, sino que solo son estados en su relación con otros cuerpos (movimiento relativo) o en relación con el espacio absoluto (movimiento absoluto). Sin embargo, sustituir el movimiento relativo por uno que se da en un espacio geométrico implica que el nuevo concepto de movimiento deje de ser empírico para ser estrictamente matemático e hipostático (Koyré 1965 10). En definitiva, el cambio semántico y categorial del concepto de movimiento implica necesariamente una redefinición del concepto de tiempo, por no decir que una teoría sobre el lugar no es suficiente y es necesaria una teoría sobre el espacio. El lugar, en la teoría de Newton, es solamente la región del espacio ocupada por un cuerpo y lo es en cuanto es llenada por la materia, es decir, es relativa a un cuerpo. Por tanto,

mientras que, como se ha visto, en la teoría de Aristóteles la noción de lugar es determinante y existe independientemente de que sea ocupada por un cuerpo, en la teoría de Newton es completamente relativizada.

4.2. FUERZA Y MASA: DOS CONCEPTOS QUE SE APRENDEN JUNTOS

En *Mundos posibles en la historia de la ciencia*, Kuhn (2002c [1989]) explica cómo los conceptos newtonianos de fuerza y masa están íntimamente relacionados, de modo que incluso a la hora de aprender la teoría newtoniana se entrelazan junto con la segunda ley. Esto, desde el punto de vista de Kuhn, implica una determinada posición respecto a la educación científica, de tal modo que el aprendizaje de los términos científicos, de las categorías taxonómicas y de las leyes se realiza en conjunto. En otras palabras, los jóvenes científicos, al igual que el historiador en su acercamiento a teorías antiguas, no pueden aprender los términos individualmente, tampoco las leyes. A continuación, se darán unas breves notas basadas en las aportaciones de Kuhn (2002c [1989]) y de Cohen (2004) acerca de cómo esto tiene consecuencias relevantes a nivel educativo. Para ello, se acudirá al ejemplo de las categorías “fuerza” y “masa”. Sin embargo, antes de pasar a analizar tal hecho, es menester realizar unos comentarios introductorios.

En la *Definición 1*, Newton especifica su definición de *cantidad de materia*, en adelante *masa*. Este concepto aparece en los *Principia* como muestra del claro rechazo de la noción cartesiana de *extensión* y de la galileana de *peso*. Como señala Cohen (2004), las experiencias de Halley y de Richter habían demostrado que el peso varía dependiendo de la latitud terrestre. Esto llevó a Newton a buscar un concepto invariable que estuviera más allá de lo circunstancial para que ocupara un papel central en sus leyes. Dicho con palabras aristotélicas, la masa no podía ser accidental, sino una constante que debía indicar, como señala Belkind, “la cantidad de lugar impenetrable que ocupa un cuerpo” (2007 281). Como es sabido, la masa relaciona el volumen y la densidad de un cuerpo. Entonces, ¿cómo es posible que se

mantenga constante si por motivos circunstanciales, como los cambios de presión o temperatura, la densidad varía? La densidad la podemos definir como la concentración del número de las partículas de un cuerpo. Pero que dicha concentración varíe no conlleva un aumento o disminución del número de partículas, aunque sí una variación del volumen. De esta manera, siempre que la densidad aumenta, el volumen disminuye y viceversa, manteniéndose la masa invariable. Así, la cantidad de materia, al no depender solo de la extensión o volumen, sino también de la densidad, se mantiene constante, lo cual es esencial para la definición de *cantidad de movimiento* y para la segunda ley.

Por otro lado, en cuanto al concepto de fuerza, hay que señalar varias cosas. En primer lugar, introduce la inercia como una *vis insita* de la materia (*Definición III*). Es preciso subrayar que la fuerza de la inercia no es una fuerza acelerativa ni una fuerza general en el sentido estricto de la dinámica (Cohen 2004 58). Esta noción, nada novedosa en la época de Newton, venía a solventar la dificultad originada con la afirmación de que si una fuerza deja de aplicarse a un cuerpo este volverá a su lugar natural, generando problemas difíciles de resolver como el famoso caso de los proyectiles. Kepler tomó el concepto de inercia y lo atribuyó a una cualidad de la materia, como un poder interno que implica que la materia se resiste a moverse. Newton, bajo el legado de Descartes y Galileo, toma el concepto de inercia como la tendencia de un cuerpo a mantener su estado, sea este de movimiento uniforme rectilíneo o de reposo. Como la *vis insita* es una capacidad de los cuerpos, es proporcional a la masa y, por lo tanto, no puede entenderse sin esta última.

En segundo lugar, la fuerza impresa (*Definición IV*) se define como “la acción ejercida sobre un cuerpo de tal forma que le obliga a cambiar su estado de reposo o movimiento”. Pues bien, este acto ocurre solo y únicamente mientras la fuerza se está produciendo, es decir, la fuerza no permanece en el cuerpo. Esto implica que es la fuerza de inercia la que hace continuar a un cuerpo en su movimiento cuando cesa la acción de la fuerza impresa. No se puede comprender, por tanto, el movimiento (por ejemplo, de un proyectil) si no se tienen en cuenta estos dos tipos de fuerza. Ahora bien, lo más controvertido del concepto de fuerza newtoniano es el caso particular

de la fuerza centrípeta. Un ejemplo de la fuerza centrípeta es el de la gravedad¹⁷ o el del movimiento curvo de los planetas. En estos casos la acción no es perceptible y se aleja de la regla, arraigada por aquel entonces, de que la fuerza se produce por la materia en contacto con la materia y no a distancia, lo que conllevó el rechazo de esta noción por parte de pensadores como Leibniz o Huygens. Esto suponía, a su vez, dar la estocada final a la distinción entre movimiento natural y movimiento forzado. Para Newton, tanto el lanzamiento de una piedra al aire como su caída al suelo son movimientos forzados.

Una vez detallado lo anterior resultará más fácil entender el análisis de Kuhn acerca de los conceptos de masa y fuerza. Estos términos, señala el autor americano, solo pueden aprenderse juntos y a través de la teoría. Es pertinente resaltar que la inmersión en un determinado léxico científico (no en referencia al aprendizaje a modo del historiador, sino el de los estudiantes que entran en el paradigma) requiere de la adquisición de los elementos teóricos a través de la práctica científica y de los ejemplos. Por ello, aunque Newton estableciera una serie de definiciones al comienzo de su obra, es precisamente en las leyes y sus demostraciones donde se aprecian las relaciones intrínsecas entre los diferentes conceptos. Por ejemplo, los términos fuerza y masa adquieren su verdadero significado una vez que son expuestos en la segunda ley, la cual se aprende a la vez que los propios términos.

La primera ley permite distinguir cuáles son los movimientos no forzados, o sea los que no requieren ninguna fuerza. En realidad, según el físico estadounidense, la primera ley es una “consecuencia lógica de la segunda” y el hecho de que no aparezcan juntas es, para él, una estrategia pedagógica para explicar el uso de fuerza y masa por separado (Kuhn 2002c [1989] 88). Sin embargo, esto no parece convincente pues olvida las razones contextuales e históricas para tal separación. Es obvio que la primera ley podría subsumirse como un caso específico de la segunda: al no haber

¹⁷ Mientras que Aristóteles necesitaba las fuerzas de gravedad y ligereza, que no tienen ninguna causa externa, para explicar los movimientos naturales hacia arriba y hacia abajo, Newton lo simplifica y solo necesita el de gravedad.

aceleración (al ser esta cero) no hay fuerza y viceversa, con lo cual solo podría haber movimiento sin aceleración o reposo. Sin embargo, la introducción de la fuerza de la inercia y, por lo tanto, de un movimiento sin fuerza motriz era tan importante y contrario a la “creencia común” del momento que resultaba imprescindible “una declaración de un nuevo principio importante de movimiento, demasiado importante para ser un caso especial de otra ley” (Cohen 2004 69). Por lo tanto, más que una razón pedagógica, suponía una toma de postura y remarcar un cambio radical. Por otro lado, la formulación de las dos primeras leyes sigue la estela de la estructura de las formulaciones de Galileo, Huygens y Descartes. Por ejemplo, Huygens describió una primera ley para el movimiento inercial en ausencia de las fuerzas de gravedad y de resistencia del aire, y, separadamente, una segunda ley para los movimientos producidos por la acción de una fuerza: la gravedad (Cohen 2004 70).

Algo que es necesario destacar es que los conceptos de masa, peso y fuerza dejan de ser características cualitativas para ser estrictamente cuantitativos, y es precisamente en la cuantificación de estos donde se revela su interrelación. El concepto cuantitativo de fuerza requiere del uso de una balanza de resorte o de muelle —pues la de brazos solo puede medir la masa a través del equilibrio de pesos—, que está basada en la ley de Hooke¹⁸. Pero, además, también necesita de la tercera ley de Newton: “la fuerza ejercida por un peso sobre el resorte es igual y opuesta a la ejercida por el resorte sobre el cuerpo” (Kuhn 2002c [1989] 90). Es decir, el término *fuerza*, que adquiere su significado a partir de la primera y la segunda ley, necesita de la tercera para cuantificarse.

Se puede entonces describir una de las rutas, pero no la única, expuesta por Kuhn para el aprendizaje de estos términos, que muestra un claro holismo e interrelación en la teoría física newtoniana. Primeramente, se necesita entender el significado de fuerza, que no puede aprenderse si no es con las dos primeras leyes, y no puede cuantificarse sin usar la tercera y la ley de Hooke. Tras esto, se aprende

¹⁸ La ley de Hooke dice así: la fuerza ejercida sobre el resorte, como es el peso de un cuerpo, es proporcional a la longitud (elongación) que adquiere el muelle.

de forma conjunta el concepto de masa y la segunda ley de Newton, y se entenderá que “la masa de un cuerpo es proporcional a su aceleración bajo la influencia de una fuerza” (Kuhn 2002c [1989] 91), lo cual demuestra que está totalmente interrelacionada con el término *fuerza*, el cual, a su vez, en su definición introduce el concepto de aceleración. Por último, el concepto de peso se aprendería como una “propiedad relacional” entre la masa de dos o más cuerpos, para lo cual se necesita de la primera ley y del concepto de fuerza para que se entienda su carácter forzado. Esto es una pequeña muestra de cómo no se pueden entender las leyes como enunciados separados, y cómo los conceptos mutan en su inserción dentro de leyes y de la red teórica en general. Lo que se aprende empíricamente a través de la práctica (como la medición) y lo que se aprende a través del léxico y sus relaciones conforman un conjunto indivisible (Kuhn 2002c [1989] 92).

En realidad, la gran interrelación que existe entre los conceptos de masa y fuerza remite en último término a otra noción central de la teoría y la metafísica newtonianas: el concepto de materia. Ya se ha dicho que en la teoría del científico británico el concepto de materia sufre un importante giro con respecto a la física aristotélica, donde la materia está completa y pasivamente doblegada a la forma. Y es que la concepción de la materia que tenía Newton tenía implicaciones muy importantes. Desde luego, la masa, en cuanto que es la “cantidad de materia” de un cuerpo, depende de manera evidente de su perspectiva atomista acerca de la materia. Sin embargo, también tiene consecuencias para la definición del concepto de fuerza. Ya se ha señalado que Newton provocó cierta conmoción con la idea de que las fuerzas pueden actuar a distancia. Esto no tenía por qué derivarse de su mirada corpuscular de la materia ya que, por ejemplo, Descartes también era atomista, pero consideraba necesario el contacto entre partículas para que una fuerza actúe y se produzca el movimiento. De hecho, autores como Descartes o Boyle “habían proscrito las cualidades ocultas de la filosofía mecánica” (Cala 2006 168). En cambio, el atomismo de Newton era muy distinto al de Descartes. Como se puede apreciar en la Óptica, en concreto en la sección de “Queries”, Newton (1952 [1730] 376) consideraba que las partículas que componen los cuerpos “actúan a distancia” entre ellas “para producir una gran parte de los fenómenos de la naturaleza”. Esto también es patente en la *Pro-*

posición 23 del Libro II de los Principia y en su Scholium cuando habla de las fuerzas de repulsión y atracción entre partículas (Henry 2023 234).

Con todo, posiblemente sea en la *Query 31 del Libro II de la Óptica* donde sea más explícita su posición acerca de la estructura de la materia. Allí plantea que dicha estructura depende de “fuerzas de atracción” a distancia entre partículas “imperceptibles”. En esa misma sección también Newton subraya que considera que existe una homogeneidad entre las partículas últimas de la materia, lo que él denomina “homogeneidad inercial de la materia”. La química de Newton sería, por tanto, contraria a aquella que se desarrollaría después basada en la idea de que existen distintos elementos químicos de diferente composición y naturaleza. Para Newton, entonces, todo cuerpo esta conformado por partículas semejantes. Sin embargo, esto le llevaba a otra problemática, como explica Cala:

Ahora bien, la homogeneidad inercial también planteaba dificultades para una consistente teoría sobre la composición de la materia, porque si se repara en que Newton ha postulado, desde los Principia, la proporcionalidad entre peso y masa inercial, ¿cómo explicar entonces la diferencia de peso entre dos cuerpos de igual volumen si sus componentes tienen la misma densidad? (2006 171).

La única respuesta posible a tal pregunta es la de la inclusión del vacío en la estructura de la materia. Es decir, Newton necesitaba suponer el vacío intramaterial. De esta manera, dos cuerpos con el mismo volumen pueden pesar distinto en la Tierra porque tienen distinta “cantidad de materia” (masa), es decir, distinta proporción de partículas y vacío. Por lo tanto, el postulado de que las partículas elementales que componen los cuerpos son homogéneas le lleva a la necesidad de suponer el vacío intramaterial, y esto, a su vez, el requerimiento de un concepto de fuerza a distancia también a nivel corpuscular. La interrelación entre estos tres conceptos fundamentales —materia, masa y fuerza— se asienta en las bases de su doctrina y conlleva una categorización concreta del mundo físico incompatible y difícilmente concebible para un aristotélico. Como se puede apreciar, la introducción por parte de Newton

de los distintos elementos teóricos analizados no puede realizarse separadamente y parece confirmar la tesis kuhniana del holismo semántico y teórico a nivel local en las teorías científicas.

5. KUHN Y LA INCONMENSURABILIDAD TAXONÓMICA

Es muy posible que el poder de la obra de Newton no resida solo en el cambio conceptual, sino también en fundar una praxis científica, un modo de hacer ciencia que pudiera abarcar tales conceptos. Lo que Cohen denomina el *estilo newtoniano* no puede desligarse de su revolución conceptual. Sin la matematización de la ciencia y cierta desontologización del cosmos, como rasgos distintivos que proponía Koyré, no se podría haber comprendido, por ejemplo, el espacio absoluto ni el papel que ha de ocupar en tal teoría. Será, por ello, que Newton en su famoso *Scholium* propone “destruir” las concepciones convencionales ligadas a las nociones o a las magnitudes de tiempo, lugar, movimiento y espacio para, así, derruir los “prejuicios” asociados a estas (Newton 1974 [1687] 6). Conseguir esto fue el culmen de una revolución que sus sucesores consumaron.

Parece que, en el caso aquí expuesto, la teoría kuhniana sobre la inconmensurabilidad taxonómica se cumple. Como se ha intentado mostrar con el precedente análisis, tanto en la física aristotélica como en la teoría newtoniana hay una serie de conceptos centrales que se interdefinen y delimitan sus implicaciones en conjunto, esto es, holísticamente, de manera que “un cambio en el significado de una parte de la estructura léxica dará como resultado un cambio en todas sus partes” (Bird 2022). Por lo tanto, se podría concluir que no se pueden entender las teorías aristotélica y newtoniana sin aprender en conjunto los significados y las implicaciones concretos de sus categorías taxonómicas centrales. Esto se debe a que tales categorías imponen al mundo estructuras taxonómicas distintas, es decir, una división en clases naturales irreconciliable (Kuhn 1993b). Intentar acomodar una teoría en la estructura taxonómica de la otra violaría lo que Kuhn (2002d 115) denomina “principio de

no-solapamiento”. Como se puede observar, de estas ideas se desvela una mirada no realista acerca de las clases naturales y las categorías, al depender estas del paradigma.

Irremediablemente, la cuestión que a este artículo ocupa desemboca finalmente en el antiguo problema de los universales. Ya Ian Hacking (1993) señalaba que la afirmación kuhniana de que tras una revolución los científicos “viven en mundos distintos” se reduce al clásico tema de los universales¹⁹. Un nominalista, subraya Hacking (1993 277), considera que, como solo existen los individuos, estos no se modifican con el cambio paradigmático. Son las descripciones las que varían, las cuales necesariamente acuden a universales, porque el mundo en el que trabajan los científicos es un mundo de clases naturales. De hecho, toda descripción, afirma el filósofo canadiense, requiere clasificar, “agrupar a los individuos en clases”. Estas descripciones y sus inherentes clasificaciones son las que cambiarían con una revolución, y no los individuos.

Como decía Nietzsche (1996 [1873] 33), “la ciencia trabaja inconteniblemente en ese gran *columbarium* de los conceptos, necrópolis de las intuiciones”, y una revolución científica significa derrumbar los cimientos de ese *columbarium* de conceptos y construir otro en su lugar, construir otro mundo fenoménico (Hoyningen-Huene 1993 33-ss.) a través de la reelaboración de las clases y los géneros de las clasificaciones científicas. Sin embargo, Kuhn, a diferencia de Kant, considera que las categorías no están dadas de antemano, no nacen aisladas e independientes de lo experiencial:

[...] la posición que estoy desarrollando es una especie de kantismo posdarwiniano. Como las categorías kantianas, el léxico proporciona las condiciones previas de las experiencias posibles. Pero las categorías léxicas, a diferencia de

¹⁹ Podría ser interesante discutir si la teoría de Kuhn implica una postura nominalista, como afirma McDonough (2003) o, en cambio, se acerca más a un conceptualismo. En este artículo no hay cabida para una indagación sobre este tema, pero es una cuestión digna de análisis para trabajos posteriores.

sus antepasadas kantianas, pueden cambiar y lo hacen, tanto con el tiempo como con el paso de una comunidad a otra. (Kuhn 2002d 129)

Desde luego, al echar por tierra la objetividad de las categorías kantianas, Kuhn irremediablemente desemboca en un rechazo del realismo alético: “usualmente se ha sostenido que las ciencias naturales [...] son inmunes. Que sus verdades (y falsedades) trascienden los estragos del cambio temporal, cultural y lingüístico. Naturalmente, estoy sugiriendo que no pueden hacerlo” (Kuhn 2002c [1989] 96). Para el pensador americano, de la misma manera que los seres vivos tienen diferentes adaptaciones que solo son útiles o válidas en determinados entornos y circunstancias, las estructuras léxicas son más o menos “adecuadas para algunos propósitos”. De manera análoga, cada lenguaje, que implica un modo de categorización, conlleva una “forma de vida”, cuyo valor es fundamentalmente instrumental:

La experiencia y la descripción solo son posibles con el descriptor y lo descrito separados, y la estructura léxica que marca la separación puede hacerlo de distintos modos, cada uno de los cuales da como resultado una forma de vida diferente, aunque nunca totalmente diferente. Algunos de dichos modos son más adecuados para algunos propósitos [...]. Pero ninguno ha de ser aceptado como verdadero o rechazado como falso; ninguno da un acceso privilegiado a un mundo real, en contraste con uno inventado. Los modos de estar-en-el-mundo que proporciona un léxico no son candidatos a verdadero/falso. (Kuhn 2002d 129)

De la anterior cita hay dos cuestiones que me gustaría destacar. En primer lugar, ese “aunque nunca totalmente” es lo que deja espacio a la comparación entre teorías. En otras palabras, que ese holismo sea local y que la mutación de significado afecte solo o especialmente a determinados términos centrales posibilita que el historiador pueda utilizar aquello que se mantiene para aprender la estructura léxica ajena. Kuhn parece tomar el principio aristotélico de que *en todo cambio siempre hay algo que permanece* cuando afirma que “en todos estos procesos de diferenciación y

cambio debe haber algo que sea permanente, fijo y estable” (Kuhn 2002d 129). Y es que para Kuhn (2002c [1989] 98) gran parte del vocabulario se mantiene, además, ocupando las mismas “posiciones” en la estructura léxica. Este es un aspecto esencial que permite el aprendizaje y la comparación, y que lo aleja del irracionalismo y radicalismo al que muchos le condenaron.

La segunda cuestión que reseñar del párrafo citado es el instrumentalismo que parece desvelar. En *Aftewords*, Kuhn (1993b) es claro a este respecto: aprender un léxico implica la adquisición de una capacidad de reconocimiento de lo estable, de lo común a los distintos fenómenos individuales que subsumimos bajo una misma clase. De esta manera, aprender el léxico de la teoría aristotélica implica ser capaz de desenvolverse con él, es decir, poder reconocer, clasificar y aplicar de manera adecuada los términos centrales de la teoría. Así, aprender el léxico aristotélico conlleva ser capaz de ordenar el mundo según la estructuración taxonómica que dicta: por ejemplo, reconocer a una planta creciendo como un movimiento o partir de la base de que todo objeto natural tiende al movimiento de actualización hacia su *telos* y que esto implica una concepción concreta y no pasiva del “lugar”, en vez de la asunción de la actuación de fuerzas a distancia. Como explica Mayoral de Lucas (2017 125), entrar en un paradigma implica adquirir un léxico científico, y esto no es otra cosa que saber aplicarlo o, como diría Wittgenstein, “saber cómo seguir la regla”, es decir, obtener un “método de reconocimiento”, un modo con el cual “identificar nuevos ejemplares de una clase”, tal y como esa teoría propone. En otras palabras, es aprender a circunscribir los fenómenos en las categorías disponibles por una teoría, y, así, ser capaz de dividir, estructurar y explicar el mundo acorde a dichas taxonomías.

Es esta capacidad práctica que otorga un léxico científico la que da acceso a un mundo fenoménico exclusivo. Por ello, es imposible la comunicación entre dos científicos monolingües pues solo tienen acceso a sus respectivos mundos fenoménicos. Mientras no se haga el esfuerzo hermenéutico de aprender el lenguaje de la otra teoría, el mundo que su estructura léxica ordena es inaccesible. En consecuencia, la incomunicabilidad entre científicos en distintos paradigmas es fruto de la infabilidad: no se puede expresar la teoría aristotélica en lenguaje newtoniano. La estructura léxica de la teoría de Newton no da acceso al mundo fenoménico estructurado por la

teoría aristotélica. Ante esto no sirven intentos de aprendizaje aislado de conceptos, es menester educarse en el conjunto, en las interconexiones léxicas y conceptuales y en desarrollar la capacidad de reconocimiento taxonómico.

6. CONCLUSIONES

Si se atiende a todo lo expuesto, una de las conclusiones principales es que, como señala Kuhn, no se puede acceder a la física aristotélica desde el léxico y los conceptos de la teoría newtoniana y viceversa. Relacionado con esto, se puede observar que dichos términos son intraducibles porque dependen de relaciones conceptuales holistas e intrincadas. Por consiguiente, la manera de acercarse a ambas teorías y de poder compararlas es aprendiendo todos esos conceptos en su interrelación, familiarizarse con el léxico específico, contextualizándolo y comprendiendo su sentido y coherencia dentro del sistema teórico. Aun en este sentido más local de inconmensurabilidad, se podría decir que no hay medida común o lenguaje neutral desde el cual realizar la comparación. La única manera es realizar un esfuerzo hermenéutico e historiográfico, siempre imperfecto, que permita comprender la transición de un paradigma hasta el punto de desarrollar la capacidad pragmática de ordenar, separar y clasificar los fenómenos y objetos de una manera coherente con la teoría aprendida. Este esfuerzo, sin embargo, rara vez lo hace un científico inmerso en su sistema conceptual con respecto a su paradigma anterior. De esta forma, la comunicación entre paradigmas parece quedar como algo prácticamente exclusivo del historiador.

La física como ciencia particular siempre fue la base y el territorio de donde Kuhn podía extraer los ejemplos que más se amoldaban a su teoría acerca de la estructura y el desarrollo científicos. Si esto es extensible a toda ciencia y a toda circunstancia es una cuestión ampliamente discutida y difícilmente resoluble. En la física hicieron falta veintiún siglos para la revolución newtoniana y, sin embargo, menos de tres para la revolución que tuvo en Albert Einstein a su máximo artífice. Es difícil estructurar la historia como si esta tuviera un sentido independiente al que se le impone. Posiblemente, el mayor pecado de Kuhn sea que, tras luchar contra las interpretaciones whig de la historia, él mismo intentó dar una estructura homogénea al desarrollo histórico de la ciencia, donde el devenir científico y el cambio para-

digmático están marcados siempre por una serie de fases sobradamente conocidas y expuestas en su obra madre. Al igual que el juez que comete el propio delito que condena, poner grilletes fijos a la historia y construir una filosofía e historia estructural de la ciencia con base en los ejemplos de un solo campo puede considerarse, como mínimo, una generalización injustificada. Como expresa Ian Hacking (2017 12), “Kuhn tomó como modelos a la física y su historia. [...] hasta qué punto lo que dijo sobre esa ciencia conserva vigencia en el abarrotado mundo biotecnológico de hoy” es digno de controversia.

La discontinuidad histórica que tanto promulgaba Kuhn (y también Koyré) puede que no sea tanto una característica de la historia como de quién la mira. La discontinuidad se da desde los ojos del newtoniano al intentar acudir a Aristóteles porque lo hace a través de las herramientas proporcionadas por la teoría en la que se ha educado y sin comprender lo intrínsecamente ligados que están los conceptos y enunciados de la teoría antigua. La discontinuidad podría ser algo propio de la educación y no tanto un hecho histórico disociador de etapas. Se puede decir que con la revolución newtoniana hay una ruptura en los términos expuestos, pero, por supuesto, jamás podrá considerarse un hecho aislado, inmediato o no causado por sus antecedentes, algo con lo que creo que estaría de acuerdo Kuhn (cf. 2002c [1989] 111). Con la obra de Copérnico y la asimilación de esta comienza el desmantelamiento de la física aristotélica, lo cual implicaba muchísimo más que un resquebrajamiento científico, pues también lo era religioso y filosófico; era atacar creencias arraigadas en la sociedad y era arremeter contra prácticas como las astrológicas, que se veían justificadas o reforzadas por el universo aristotélico. Solo a partir de entonces, como piezas de un mismo rompecabezas, empiezan a fraguarse algunos de los elementos que posteriormente permitirían a Newton convertirse en el actor central de una revolución que comenzó un siglo antes de su nacimiento. Por esta razón, más allá de que la conclusión de este artículo sea que existe esa inconmensurabilidad taxonómica entre la física aristotélica y la newtoniana, también se quiere resaltar que no hay forma de justificar la afirmación de Kuhn (2002b [1987] 74) de que “los cambios holísticos tienden a ocurrir de repente”.

REFERENCIAS

- Achtner, Wolfgang, Kunz, Stefan y Walter, Thomas. *Dimensions of Time. The Structures of the Time of Humans, of the World, and of God*. Cambridge: Eerdmans, 2012.
- Aristóteles. “Física”. *Aristóteles II*. Madrid: Gredos, 2011.
- _____. *Acerca del cielo*. Madrid: Gredos, 1996.
- Aubenque, Pierre. *El problema del ser en Aristóteles*. Madrid: Taurus, 1981.
- Belkind, Ori. “Newton's conceptual argument for absolute space”. *International Studies in the Philosophy of Science* 21.3 (2007): 271-293. <<https://doi.org/10.1080/02698590701589551>>
- _____. “Leibniz and Newton on space”. *Foundations of Science* 18.3 (2013): 467-497. <<https://doi.org/10.1007/s10699-011-9280-5>>
- _____. “Action, substance, and space in Newton's metaphysics”. *Iyyun: The Jerusalem Philosophical Quarterly* 69 (2021): 31-56. <https://www.researchgate.net/publication/356459218_Action_Substance_and_Space_in_Newton's_Metaphysics>
- Bird, Alexander. “Thomas Kuhn”. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.), 2022. <<https://plato.stanford.edu/archives/spr2022/entries/thomas-kuhn/>>
- Cadavid-Ramírez, Lina M. y Arias-Vélez, Mateo. “La inconmensurabilidad kuhniana y las generalizaciones simbólicas: un análisis a partir de la comparación de dos paradigmas teóricos de la física”. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales* 45.175 (2021): 377-386. <<https://doi.org/10.18257/raccefyn.1258>>
- Cala Vitery, Favio Ernesto. “La cuestión 31 de la óptica o el programa de las fuerzas en la filosofía mecánica”. *Scientiae Studia* 4.2 (2006): 163-76. <<https://doi.org/10.1590/S1678-31662006000200002>>
- Casini, Paolo. *El universo máquina*. Barcelona: Martínez Roca, 1971.

- Chen, Chung-Hwan. The relation between the terms 'energeia' and 'entelecheia' in the philosophy of Aristotle. *Classical Quarterly* 8.1-2 (2009): 12-17. <<https://doi.org/10.1017/S0009838800016360>>
- Cohen, I. Bernard. "Newton's concepts of force and mass, with notes on the laws of motion". *The Cambridge Companion to Newton*. I. B. Cohen y G. E. Smith (eds.). Cambridge University Press, 2004. 57-84. <<https://doi.org/10.1017/CCOL0521651778.003>>
- Conill, Jesús. *El tiempo en la filosofía de Aristóteles. Un estudio dedicado especialmente al análisis del Tratado del tiempo (Física IV, 10-14)*. Valencia: Facultad de Teología San Vicente Ferrer, 1981.
- Descartes, René. *El mundo o tratado de la luz*. México D. F.: UNAM, 1986 [1664].
- DiSalle, Robert. "Newton's Philosophical Analysis of Space and Time". *The Cambridge Companion to Newton*. Robert Iliffe y George E. Smith (eds.). Cambridge University Press, 2004. 33-56. <<https://doi.org/10.1017/CCOL0521651778.002>>
- _____. "Absolute space and Newton's theory of relativity". *Studies in History and Philosophy of Science Part B* 71 (2020): 232-244. <<https://doi.org/10.1016/j.shpsb.2020.04.003>>
- Einstein, Albert. "Prólogo". *Conceptos de espacio*. Max Jammer (ed.). Ciudad de México: Grijalbo, 1970. 11-17.
- Friedman, Michael. *Foundations of Space-Time Theories*. Princeton University Press, 1983.
- _____. "Newton and Kant on absolute space: From theology to transcendental philosophy". *Interpreting Newton. Critical Essays*. Andrew Janiak y Eric Schliesser (eds.). Cambridge University Press, 2012. 342-359. <<https://doi.org/10.1017/CBO9780511994845.018>>
- Hacking, Ian. "Working in a new world: The taxonomic solution". *World Changes: Thomas Kuhn and the Nature of Science*. Paul Horwich (ed.). Cambridge/Massachusetts/Londres: MIT Press, 1993. 275-310. <<https://doi.org/10.2307/jj.11374771.13>>

- _____. “Ensayo preliminar”. *La estructura de las revoluciones científicas*. Thomas S. Kuhn. Madrid: Fondo de Cultura Económica, 2017.
- Hall, A. Rupert. “Newton and the absolute sources”. *The Investigation of Difficult-Things Essays on Newton and the History of the Exact Sciences*. Peter M. Harman y Alan E. Shapiro (eds.). Cambridge University Press, 1992. 261-285.
- Henry, John. “Newton’s ‘De aere et aethere’ and the introduction of interparticulate forces into his physics”. *Annals of Science* 80.3 (2023): 232-267. <<https://doi.org/10.1080/00033790.2023.2192721>>
- Hoyningen-Huene, George. *Reconstructing Scientific Revolutions*. The University of Chicago Press, 1993.
- _____. “Kuhn’s development before and after structure”. *Kuhn’s Structure of Scientific Revolutions: 50 Years On*. William J. Devlin y Alisa Bokulich (eds.). Springer, 2015. 185-195
- Jammer, Max. *Conceptos de espacio*. Ciudad de México: Grijalbo, 1970.
- Janiak, Andrew. *Newton as philosopher*. Cambridge University Press, 2008. <https://doi.org/10.1017/CBO9780511481512>
- _____. “Substance and action in Descartes and Newton”. *The Monist* 93.4 (2010): 657-677. <<https://doi.org/10.5840/monist201093437>>
- _____. “Metaphysics and natural philosophy in Descartes and Newton”. *Foundations of Science* 18 (2013): 403-417. <<https://doi.org/10.1007/s10699-011-9277-0>>
- Jarnicki, Paweł y Greif, Hajo. “The ‘Aristotle experience’ revisited: Thomas Kuhn meets Ludwik Fleck on the road to structure”. *Archiv für Geschichte der Philosophie* (2022). <<https://doi.org/10.1515/agph-2020-0160>>
- Koyré, Alexander. *Newtonian Studies*. Londres: Chapman and Hall, 1965.
- _____. *Estudios galileanos*. Madrid: Siglo XXI Editores, 1980.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*. University of Chicago Press, 1970.
- _____. *La tensión esencial. Estudios selectos sobre la tradición y el cambio en el ámbito de la ciencia*. Madrid: Fondo de Cultura Económica, 1993a.

- _____. “Afterwords”. *World Changes Thomas Kuhn and the Nature of Science*. Ed. Paul Horwich. Cambridge/Massachusetts/Londres: MIT Press, 1993b. 311-341.
- _____. *La revolución copernicana. La astronomía planetaria en el desarrollo del pensamiento*. Barcelona: Ariel, 1996.
- _____. “Conmensurabilidad, comparabilidad y comunicabilidad”. *El camino desde la estructura*. James Conant y John Haugeland (comps.). Barcelona: Paidós, 2002a [1982]. 47-75.
- _____. “¿Qué son las revoluciones científicas?”. *El camino desde la estructura*. James Conant y John Haugeland (comps.). Barcelona: Paidós, 2002b [1987]. 23-45.
- _____. “Mundos posibles en la historia de la ciencia”. *El camino desde la estructura*. James Conant y John Haugeland (comps.). Barcelona: Paidós, 2002c [1989]. 76-112.
- _____. “El camino desde *La estructura*”. *El camino desde la estructura*. James Conant y John Haugeland (comps.). Barcelona: Paidós, 2002d. 113-129.
- _____. “Una conversación con Thomas Kuhn”. *El camino desde la estructura*. James Conant y John Haugeland (comps.). Barcelona: Paidós, 2002e. 299-374.
- Kuukkanen, Jouni-Matti. *Meaning Changes: A Study of Thomas Kuhn's Philosophy*. AkademikerVerlag, 2012.
- Mayoral de Lucas, Juan Vicente. “Mundos fenoménicos y léxicos científicos: el relativismo lingüístico de Thomas Kuhn”. *Revista de Filosofía* 42.1 (2017): 117-134. <<https://doi.org/10.5209/RESF.55452>>
- McDonough, Jeffrey K. “A ‘Rosa multiflora’ by any other name: Taxonomic incommensurability and scientific kinds”. *Synthese* 136 (2003): 337-358. <<https://doi.org/10.1023/A:1025116814353>>
- McGuire, James E. y Tanmy, Martin. *Certain Philosophical Questions: Newton's Trinity Notebook*. Cambridge University Press, 1983.
- Musso, Paolo. “Why Newton's absolute space-time is not so absolute and Einstein's relative space-time is not so relative”. *Epistemologia* 37.1 (2014): 152-157. <<https://doi.org/10.3280/EPIS2014-001010>>

- Newton, Isaac. *Opticks*. Nueva York: Dover Publications, 1952 [1730].
- _____. *Mathematical Principles of Natural Philosophy*. Vols. I y II. Berkeley: University of California Press, 1974 [1687].
- _____. “De gravitatione”. *Newton. Philosophical Writings*. Andrew Janiak (ed.). Cambridge University Press, 2004. 12-39.
- Nietzsche, Friedrich. *Sobre verdad y mentira en sentido extramoral*. Madrid: Tecnos, 1996 [1873].
- Qu, Li. “Newton, Einstein and Barth on Time and Eternity”. *Scottish Journal of Theology* 67 (2014): 436-449. <<https://doi.org/10.1017/S0036930614000209>>
- Reisch, George A. “Aristotle in the Cold War: On the origins of Thomas Kuhn’s structure of scientific revolutions”. *Kuhn’s ‘Structure of Scientific Revolutions’ at Fifty*. Robert J. Richards y Lorraine Daston (eds.). University of Chicago Press, 2016. 12-30. <<https://doi.org/10.7208/chicago/9780226317175.003.0001>>
- Rorty, Richard. *La filosofía y el espejo de la naturaleza*. Madrid: Cátedra, 1983.
- Ross, William D. *Aristóteles*. Buenos Aires: Sudamericana, 1957.
- Rossi, Paolo. *El nacimiento de la ciencia moderna en Europa*. Barcelona: Crítica, 1998.
- Rynasiewicz, Robert. “Newton’s views on space, time, and motion”. *The Stanford Encyclopedia of Philosophy*. Edward Zalta (ed.), 2014. <<https://plato.stanford.edu/archives/sum2014/entries/newton-stm/>>
- Schliesser, Eric S. “Newton and Spinoza: On motion and matter (and God, of course)”. *The Southern Journal of Philosophy* 50.3 (2012): 436-458. <<https://doi.org/10.1111/j.2041-6962.2012.00132.x>>
- _____. “Newtonian emanation, Spinozism, measurement, and the Baconian origins of the laws of nature”. *Foundations of Science* 18 (2013): 449-466. <<https://doi.org/10.1007/s10699-011-9279-y>>
- Sellés, Manuel y Solís, Carlos. *Revolución científica*. Madrid: Síntesis, 1994.
- Stein, Howard. “Newton’s metaphysics”. *Cambridge Companion to Isaac Newton*. I. Bernard Cohen and George E. Smith (eds.). Cambridge University Press, 2002.
- Strong, Edward W. “Barrow and Newton”. *Journal of the History of Philosophy* 8.2 (1970): 155-172. <<https://doi.org/10.1353/hph.2008.1726>>

- Voltaire. François-Marie. “Cartas filosóficas”. *Voltaire 1*. Madrid: Gredos, 2010 [1734]. 1-161.
- Van Dyck, Maarten y Verelst, Karin. “‘Whatever is neither everywhere nor anywhere does not exist’: The concepts of space and time in Newton and Leibniz”. *Foundations of Science* 18 (2013): 399-402. <<https://doi.org/10.1007/s10699-011-9276-1>>
- Van Fraassen, Bas C. *An Introduction to the Philosophy of Time and Space*. Nueva York: Random House, 1970.
- Wray, Brad. *Kuhn’s Intellectual Path*. Cambridge University Press, 2021.

INDICACIONES PARA LOS AUTORES

La *Revista Colombiana de Filosofía de la Ciencia* es una publicación académica dedicada a la filosofía de la ciencia y a sus campos afines (lógica, epistemología, ciencias cognitivas, filosofía de la tecnología, filosofía del lenguaje) y, en general, a los temas y problemas que ponen en diálogo a las ciencias con la filosofía. En ocasiones se editan números monográficos sobre autores o temas puntuales. La revista recibe contribuciones en forma de artículos originales y reseñas de libros en español, portugués, francés e inglés. Todas las colaboraciones serán evaluadas por un árbitro de manera anónima y el autor recibirá una respuesta en un lapso no mayor a 90 días. Se entiende que los autores autorizan a la revista la publicación de los textos aceptados en formato impreso y digital.

Todas las contribuciones han de ser enviadas en formato doc, docx, o rtf por correo electrónico a la dirección revistafilosofiaciencia@unbosque.edu.co, y han de cumplir con las siguientes condiciones:

ARTÍCULOS

- El texto ha de ser original e inédito y no se ha de encontrar en proceso de evaluación para su publicación por ninguna otra revista académica.
- Se ha de enviar el artículo en un archivo, en versión anónima y cuidando que las notas a pie de página, agradecimientos o referencias internas en el texto no revelen la identidad de su autor. En un archivo aparte se ha de enviar el título del artículo, el nombre del autor, su afiliación institucional y sus datos de contacto (dirección de correspondencia, correo electrónico y teléfono).
- El artículo debe venir precedido de un resumen en su idioma original que no exceda las 100 palabras, y 5 palabras clave. Se han de incluir también las traducciones al inglés del título del artículo, el resumen y las palabras clave.

- La lista de trabajos citados ha de estar al final del artículo y ha de cumplir con el sistema MLA de la citación para el área de filosofía (<http://www.mla.org/style>).
- Las referencias bibliográficas han de incorporarse al texto y no en las notas al pie de página (las notas a pie de página han de restringirse así a aquellas que contengan información sustantiva), de la siguiente manera: (Autor, página). En caso de que haya más de una obra del autor en la bibliografía, se ha de agregar el año de la obra: (Autor, año, página).
- Las citas textuales de más de cinco líneas han de ubicarse en párrafo aparte con sangría de 0,5 cms. a margen derecho e izquierdo, y no han de estar entrecomilladas. Las citas de extensión menor no requieren párrafo aparte y han de venir entrecomilladas.
- La extensión máxima de los artículos es de 15.000 palabras.

RESEÑAS BIBLIOGRÁFICAS

- Se recibirán únicamente reseñas sobre libros publicados recientemente (cuya fecha de publicación no exceda los últimos dos años).
- Las reseñas han de cumplir con las mismas condiciones para la citación, notas al pie y referencias bibliográficas ya especificadas para los artículos.
- La extensión máxima de las reseñas es 2.500 palabras.
- Los autores de artículos y reseñas que sean publicados en la revista recibirán dos ejemplares de la misma.

DERECHOS DE AUTOR
INFORMACIÓN PARA OBTENER Y REPRODUCIR
DOCUMENTOS PUBLICADOS

Al postular un artículo para su posible publicación, los autores conceden implícitamente su autorización a la Revista para publicarlo. La publicación del artículo en la Revista supone que los derechos de autor patrimoniales pasan a ser propiedad de la institución editora de la Revista, la Universidad El Bosque. La Revista ha definido un formato de cesión de derechos de autor que deberá firmar todo autor que presente su obra para ser considerada en el Comité Editorial. En este sentido, las solicitudes para reproducir artículos publicados en la Revista podrán enviarse por correo electrónico al editor, indicando la referencia completa del material que se desea emplear (volumen, número, año, autor, título del artículo, número de páginas). Es necesario, además, indicar el uso que se pretende dar al material (uso total o parcial, tipo de publicación, institución del editor, fecha aproximada de publicación). El contenido de los artículos es responsabilidad absoluta de sus autores y no compromete, en ningún caso, a la Revista o a la Universidad.

INSTRUCTIONS FOR AUTHORS

The *Revista Colombiana de Filosofía de la Ciencia* is an academic journal published by the Humanities Department of the Universidad El Bosque, mainly devoted to the Philosophy of Science and their related fields (Epistemology, Logic, Cognitive Science, Philosophy of Technology, Philosophy of Language) and, in general, the topics and problems that generate dialogue between philosophy and science, whether pure sciences, applied, social or human. Sometimes issues are published on specific topics or authors. The journal receives submissions in the form of original articles and book reviews in Spanish, Portuguese, French and English. Submissions received will be considered by the editorial committee for publication, verifying that they fit their own areas of the journal; after receipt they will be evaluated by an anonymous expert referee and the author will receive a response within a period not exceeding 90 days. It is understood that the authors authorize publication of accepted texts in print and digital.

All submissions must be sent in Word, docx or rtf format, and emailed to the address revistafilosofiaciencia@unbosque.edu.co, and they must meet the following conditions:

ARTICLES

- The text must be original, unpublished and should not be under evaluation for publication by any other journal.
- The author must send the manuscript in a file, in anonymous version and making sure that the footnotes, acknowledgments and internal references in the text does not reveal the identity of its author. In a separate file, the author must include: the article title, author's name, institutional affiliation and contact information (mailing address, email and phone).

- The paper must be preceded by a summary in the original language that does not exceed 100 words and 5 keywords. It should also include the English translations of the article title, abstract and keywords (or the Spanish translation, if the original language of the article is English).
- The complete list of works cited must be at the end of the article and must comply with the MLA citation system for the area of philosophy (<http://www.mla.org/style>).
- References must be incorporated into the text and not in footnotes (the footnotes have to be restricted to those that contain substantive information), as follows: (Author page). If there is more than one work by the same author in the bibliography, in the reference must be added the year of the work: (Author year page).
- Quotations of more than five lines must be placed in a separate paragraph indented 0.5 cm to left and right margins, and don't need quotations marks. The quotations of minor extension don't require a separate paragraph.
- The maximum length of articles is 15,000 words.

BOOK REVIEWS

- It will be received only reviews of recently published books (whose publication date must not to exceed two years).
- The review must meet the same conditions for the citation, footnotes and list of works cited for articles already specified.
- The maximum length of the reviews is 2,500 words.
- The authors of articles and reviews published in the journal will receive two copies of it.

COPYRIGHT - INFORMATION FOR DOCUMENTS ISSUED

By postulating an article for publication, the authors implicitly granted permission to the Journal for publication. The publication of the article in the Journal supposed author property rights become the property of the institution editor of the Journal, Universidad El Bosque. The Journal has defined a format copyright assignment to be signed by all authors to submit their work for consideration in the Editorial Committee. In this regard, requests to reproduce articles published in the Journal may be sent by email to the editor, indicating the complete reference material to be used (volume, number, year, author, article title, number of pages). It is also necessary to indicate the intended use give the material (total or partial use, type of publication, institution editor, and approximate date of publication). The content of the articles is the sole responsibility of the authors and not compromise under any circumstances, to the Journal or to the University.



Departamento de humanidades
Programa de filosofía

Junio de 2024
Bogotá, Colombia

Contenido

La economía folk y la ciencia económica

Luis Enrique Ortiz Gutiérrez - México

Traducción: Folk economics and economic science

Luis Enrique Ortiz Gutiérrez - México

Elementos para una revisión epistemológica de la axiomática de Kolmogorov

Alberto Landro y Mirta L. Gonzalez - Argentina

Traducción: Elements for an epistemological review of Kolmogorov axiomatics

Alberto Landro y Mirta L. Gonzalez - Argentina

El simio egoísta: críticas a la hipótesis del pensamiento maquiavélico

Mateo Arias-Vélez - México

La posibilidad de explicación científica a partir de modelos basados en redes neuronales artificiales

Alejandro E. Rodríguez-Sánchez - México

Entidades físicas, identidad temporal y basura espaciotemporal.

Una reconstrucción crítica de la ontología física de Dowe

Wilfredo Quezada Pulido y Luis Pavez - Chile

Holismo local y cambio semántico en la teoría de Kuhn: aristóteles, newton y la inconmensurabilidad taxonómica

Daniel Labrador-Montero - España

Indicaciones para los Autores



UNIVERSIDAD
EL BOSQUE

Departamento de Humanidades
Programa de Filosofía
Maestría en Filosofía de la Ciencia

Av. Cra 9 N° 131 a -02 Edificio Fundadores

Línea gratuita 018000113033 * PBX (571) 6489000

Bogotá D.C. - Colombia