

Serie  
*Selección de Textos*

S|T

# ESTUDIOS SOBRE FILOSOFÍA Y CIENCIA

---

Mario Tapia Ramírez & Lucas Hinojosa-López  
(Editores)



Serie  
*Selección de Textos*

**S | T**



# Serie Selección de Textos

Juan Redmond & Rodrigo Lopez-Orellana (Directores)  
Jorge Budrovich S. (Editor)

## Volumen 10

---

# Estudios sobre Filosofía y Ciencia

MARIO TAPIA RAMÍREZ

LUCAS HINOJOSA-LÓPEZ

(Editores)

Universidad de Valparaíso  
Facultad de Humanidades y Educación  
Instituto de Filosofía

2025

FACULTAD DE HUMANIDADES Y EDUCACIÓN  
UNIVERSIDAD DE VALPARAÍSO

Rector: Osvaldo Corrales Jorquera  
Vicerrector de Gestión Institucional: Christian Corvalán Rivera  
Secretario General: Ricardo Bravo Méndez  
Decano: Pablo Aravena Núñez

*Comité Editorial de la Serie*

Directores: Juan Redmond & Rodrigo Lopez-Orellana  
Editor: Jorge Budrovich Sáez

*Comité Científico de la Serie*

Adriana Arpini, Universidad Nacional de Cuyo, Argentina  
Alejandro Cassini, Universidad de Buenos Aires, Argentina  
Carlos Bello, UTN-Facultad Regional Mendoza, Argentina  
Claudio Albertani, Universidad Autónoma de la Ciudad de México  
Cristián Soto, Universidad de Chile, Chile  
Diego Fernandes, Universidade Federal de Goiás, Brasil  
Esteban Anzoise, UTN-Facultad Regional Mendoza, Argentina  
Felix Aguirre, Universidad de Valparaíso, Chile  
Guillermo Cuadrado, UTN-Facultad Regional Mendoza, Argentina  
Jaime Villegas, Universidad de Valparaíso, Chile  
María José Frápolli, University College London, United Kingdom  
María Manzano Arjona, Universidad de Salamanca, España  
Miguel Tornello, UTN-Facultad Regional Mendoza, Argentina  
Rolando Rebolledo, Universidad de Valparaíso, Chile  
Rubén Quiroz Avila, Universidad Nacional de San Marco, Perú  
Sara Beatriz Guardia, Universidad San Martín de Porres, Perú  
Shahid Rahman, Université de Lille 3, Francia  
Víctor Aranda, Universidad Complutense de Madrid, España

Edición:

INSTITUTO DE FILOSOFÍA, UNIVERSIDAD DE VALPARAÍSO  
Impreso en Valparaíso, Chile. Diciembre de 2025

Serrano 546, Valparaíso. Chile

ISBN 978-956-420-357-7

# Índice

Prólogo	
JUAN REDMOND . . . . .	1
Models in Medicine: the Digital Twin for Health	
VICENTE BASAURE . . . . .	3
La proyección antropocéntrica del <i>Test de Turing</i> en la cultura de la IA	
ANLLELY GARCÍA ROJAS . . . . .	17
Conspirativismo digital y demarcación científica: un acercamiento desde la etnografía virtual	
MANUEL GARCÍA Y PABLO VERDE . . . . .	33
Procesamiento predictivo y bayesianismo: una brevísima aproximación filosófica-probabilística al modelamiento predictivo de la mente	
LUCAS HINOJOSA-LÓPEZ . . . . .	49
Teorías humeanas y No humeanas: el desacuerdo con respecto a los inobservables	
KUMINAK LEFIO ZAMORANO . . . . .	61
La propuesta de Hempel para un análisis de los juicios morales a partir de los métodos de contrastación y validación científicos	
RODRIGO LÓPEZ-ORELLANA . . . . .	79
Normativismo, naturalismo y lo patológico del delirio	
PABLO LÓPEZ-SILVA Y CLAUDIO ANABALÓN . . . . .	93
Acerca del debate sobre el rol adaptativo de los delirios en la filosofía de la psiquiatría	
PABLO LÓPEZ-SILVA, CHANNEL ALARCÓN, TIARE QUIROZ, JUAN RIVEROS, MARTINA SERRAT, PABLO ÁLVAREZ, ÁLVARO CAVIERES . . . . .	107
<i>Hypotheses fingo</i> : Interpretación, anticipación e investigación de la naturaleza en los <i>Principia Mathematica</i> de Newton	
ANDREW MAYNE-NICHOLLS MERINO . . . . .	129

---

Formas y pragmatismo en la filosofía de Francis Bacon	
SOFÍA MUÑOZ NILO . . . . .	149
Conjuntos de creencias y coherentismo: posibles manipulaciones del conocimiento científico	
MIGUEL NAREDO ROJAS . . . . .	163
Reflexiones desde un enfoque biosemiótico peirceano sobre algunos procesos de mimetismo biológico.	
LORETO PANIAGUA-VALDEBENITO . . . . .	181
Algunas observaciones críticas a la noción de modelo en el Enfoque Heredado	
JUAN REDMOND . . . . .	203
Tiempo relacional: bases para una ontología científica minimalista	
FRANCO REYES AGUIRRE . . . . .	219
El empréstito como herramienta de entrecruce en <i>Maniac</i> : entre la ciencia y la literatura	
MATHIAS RIQUELME Z. . . . .	235
Física y metafísica en Descartes. Una pregunta a partir de la lectura de Clarke.	
GABRIEL OSMAN SANDOVAL . . . . .	247
Revisitando el problema de Wigner: la irrazonable eficacia de las matemáticas en las ciencias naturales	
LEONARDO SEGURA . . . . .	263
Para una interpretación lógica de la interrogatividad de las preguntas.	
MARIO TAPIA RAMÍREZ . . . . .	277



## Prólogo

El presente libro fue escrito colectivamente y ha sido publicado con el apoyo del **Centro de Estudios en Filosofía de la Ciencia, Lógica y Epistemología** (CEFILOE) de la **Universidad de Valparaíso**. Reúne artículos ordenados alfabéticamente en la línea de *Filosofía de la Ciencia*. Los temas tratados son de notable actualidad y su diversidad acusa la fructífera creatividad de las y los autoras/es.

De este modo, el CEFILOE cumple con los objetivos de (i) Promover y desarrollar investigación de excelencia en Filosofía, Lógica y Epistemología en línea con las mejores prácticas internacionales, así como la producción de resultados con impacto nacional e internacional a través de publicaciones. (ii) Crear fuertes vínculos entre el ámbito de la investigación y el de la educación de pre y postgrado. (iii) Difundir los resultados de la investigación estableciendo canales de comunicación con la sociedad civil mediante el intercambio continuo de información y concretando debates abiertos en el ámbito de los nuevos retos y tareas sociales vinculados a los resultados de la ciencia contemporánea, en los que la Filosofía, la Lógica y la Epistemología pueden contribuir a un entendimiento recíproco. (iv) Promover y llevar a cabo iniciativas encaminadas a mejorar la formación de grado y postgrado y la educación permanente en temas principalmente relacionados con la línea de Lógica y Filosofía de la Ciencia de nuestra Pedagogía en Filosofía (Universidad de Valparaíso). (v) Fomentar la cooperación con Universidades, Institutos y Centros de Investigación similares a nivel nacional e internacional a través de la convocatoria abierta que realizamos y que permitió el intercambio de académicos especializados y la difusión del conocimiento en el ámbito de las ciencias sociales y humanas.

Al mismo tiempo, el presente libro tributa al perfil de egreso de nuestras carreras a través de resultados de aprendizaje y desempeños claves, contribuyendo con el desarrollo y aseguramiento de la calidad de la educación.

JUAN REDMOND y RODRIGO LÓPEZ-ORELLANA

**Instituto de Filosofía**  
**Universidad de Valparaíso**

# Models in Medicine: the Digital Twin for Health

Vicente Basaure<sup>a</sup>

## Resumen

Este artículo explora el papel de los modelos en la medicina con un enfoque en la evolución y aplicación de modelos matemáticos y computacionales, en particular los Gemelos Digitales (GDs). Los modelos en la ciencia sirven como herramientas esenciales para representar sistemas y fenómenos complejos, lo que permite a los investigadores predecir resultados y desarrollar estrategias efectivas. Históricamente, los modelos han sido fundamentales para los avances en varios campos médicos, desde la farmacocinética hasta la propagación de enfermedades y la respuesta tumoral a ciertos tratamientos. En la medicina moderna, los modelos predictivos, incluidos los GDs, ofrecen un potencial transformador al mejorar los resultados de los pacientes a través de la medicina personalizada, optimizar la gestión sanitaria y avanzar en la investigación biomédica. Los Gemelos Digitales, que son réplicas digitales detalladas de entidades físicas, están emergiendo como herramientas críticas en el ámbito de la salud, capaces de simular desde órganos individuales hasta sistemas hospitalarios completos. A pesar de su promesa, la implementación de los GDs enfrenta desafíos como la privacidad de los datos, la integración de información y la precisión de los modelos. Superar estos obstáculos requiere colaboración entre proveedores de salud, investigadores y desarrolladores tecnológicos. A medida que este campo avanza, se espera que estos modelos cambien significativamente el futuro de la ciencia médica.

**Palabras clave:** representación, predicción, matemático, in silico, ingeniería biomédica.

---

<sup>a</sup>Universidad de Valparaíso, Viña del Mar, Chile.  
Contacto: vibaro8@gmail.com

## Abstract

This paper explores the role of models in medicine with a focus on the evolution and application of mathematical and computational models, particularly Digital Twins (DTs). Models in science serve as essential tools for representing complex systems and phenomena, enabling researchers to predict outcomes and develop effective strategies. Historically, models have been integral to advancements in various medical fields, from pharmacokinetics to disease spread and tumor response to certain treatments. In modern medicine, predictive models, including DTs, offer transformative potential by improving patient outcomes through personalized medicine, optimizing healthcare management, and enhancing biomedical research. Digital Twins, which are detailed digital replicas of physical entities, are emerging as critical tools in healthcare, capable of simulating everything from individual organs to entire hospital systems. Despite their promise, the implementation of DTs faces challenges such as data privacy, integration, and model accuracy. Overcoming these obstacles requires collaboration among healthcare providers, researchers, and technology developers. As the field advances, these models are poised to significantly reshape the future of medical science.

**Keywords:** representation, prediction, mathematical, in silico, biomedical engineering.

## 1. Introduction

The use of models in science, particularly in medicine, has revolutionized the way researchers and practitioners understand complex phenomena. From the early days of ancient medical theories such as the balance of humors to the sophisticated mathematical and computational models employed in contemporary biomedical research, models have served as indispensable tools for representing, analyzing, and predicting various aspects of the physical world. In recent years, advancements in data science, artificial intelligence (AI), and computational power have paved the way for a transformative concept in healthcare: the Digital Twin (DT). Originally developed for engineering and space missions, the digital twin has rapidly gained prominence in medicine, promising to reshape patient care, drug development, and medical research through personalized and dynamic digital replicas of real-world entities.

The potential of Digital Twins for Health (DT4H) lies in their ability to replicate not just single organs but entire physiological systems or even healthcare environments, allowing for real-time monitoring, predictive analysis, and tailored medical interventions. These advancements, however, also bring about profound methodological, episte-

mological, and ethical challenges that must be addressed to ensure the safe and effective integration of DTs into mainstream medical practice.

This work explores the historical and contemporary use of models in medical science, focusing on the emergence and application of Digital Twins. It delves into the philosophical and practical implications of modeling, the role of predictive algorithms, and the opportunities DT4H presents for improving patient outcomes while critically assessing the challenges that accompany this evolving field. Through this lens, the discussion highlights how the Digital Twin concept has the potential to redefine the future of healthcare, making personalized medicine and *in silico* trials viable components of modern medical practice.

## 2. Models in Science

The concept of models in science is rooted in model theory, which examines the interpretation of any language—formal or natural—through set-theoretic structures. Modeling a phenomenon essentially involves constructing a formal theory to describe and explain it (Hodges 2023). According to Moulines, the model-based approach in the philosophy of science gained prominence around the 1970s. This approach emphasizes models rather than propositions as the fundamental units of scientific understanding, underscoring the importance of detailed reconstructions of scientific theories.

Scientific models are designed to represent specific parts or aspects of the world, referred to as the model's target system. These models can take various forms, including physical objects, fictional objects, abstract entities, set-theoretic structures, descriptions, and equations (Frigg 2024). Despite the multiple meanings associated with the concept of models, Cassini (2016) argues that they are fundamentally idealized representations of a phenomenon or domain. While the notion of representation is still evolving, modeling remains a central activity in modern science as cognitive vehicles.

In contemporary scientific research, models often replace the direct study of reality due to practical and economic considerations. By examining a model, researchers can uncover features and derive insights about the system it represents, enabling what is termed “surrogate reasoning.” Swoyer (1991, p. 449) explains that this process involves generating hypotheses about phenomena based on a model. Redmond (2022, p. 11) expands on this concept, suggesting that surrogation should be viewed as a logical relationship wherein conclusions drawn from the

model are applicable to the target system. In essence, surrogate reasoning allows inferences made within the model to be considered valid for the system it represents.

### 3. Models in Medical History

The concept of representation and surrogate reasoning in medicine dates back to ancient times. The theory of humors, practiced by Hippocrates and Galen, serves as an early example. This theory posited that the human body—the target system—was composed of four balanced humors—the model: black bile, yellow bile, blood, and phlegmeach influenced by environmental factors such as seasons, climate, diet, and geography. Illness was thought to result from an imbalance among these humors, and treatments aimed to restore balance through dietary adjustments, purges, enemas, herbal medications, or surgical procedures such as bloodletting (Lamas 2016, p. 82).

Although the humor theory fits the presented definition of modeling, it differs from modern modeling concepts as discussed here, particularly because the focus of this work is on modeling phenomena into formal languages, specifically mathematical and computational methods.

The modern use of models in medical science began in the 20th century, driven by technological advancements. These models employ mathematical algorithms to predict outcomes in biomedicine. For example, the Pharmacokinetic/Pharmacodynamic (PK/PD) models aid in understanding drug interactions within the body, helping determine appropriate dosages (Michaelis & Menten 1913, p. 333).

Another prominent example is the SIR Model (Susceptible-Infected-Recovered), which uses differential equations to simulate disease transmission dynamics (Kermack & McKendrick 1927, p. 115). The Hodgkin-Huxley Model (1952) describes how action potentials in neurons are initiated and propagated, using mathematical equations to represent ion channel activity in nerve cells, providing insights into neurological diseases and potential treatments (Hodgkin & Huxley 1952, p. 117).

In cancer medicine, the Tumor Growth Inhibition (TGI) Model, used since 2004, predicts how a tumor will grow or shrink in response to different treatments. This model helps oncologists determine the most effective therapy by simulating the tumor's response over time (Simeoni et al. 2004, p. 1094).

Additionally, non-mathematical models, such as high-fidelity patient simulators like SimMan, are used to replicate real-life scenarios

like cardiac arrest or respiratory failure (Alinier et al. 2006, p. 359). Also, anatomical models like The Visible Human Project, a digital model of the human body providing detailed cross-sectional views of anatomical structures, are widely used in medical education.

Lastly, it is worth mentioning projects that aim to incorporate physiological features into these computational imaging databases, like the Virtual Physiological Human Initiative, with a holistic approach to medicine where the body is treated as one single, however complex, multiorgan system rather than as a collection of individual organs. This is the base of the notion of the Digital Twin (Paul 2023, pp. 8-15).

## 4. Models in Modern Medicine

The complexities and dynamic nature of modern biomedicine require methodologies capable of processing the multiple factors involved in this field's phenomena. Mathematical and computational models, particularly predictive models, address this need. Predictive models involve the use of these methods to create models that can forecast future outcomes. The application of these techniques to the medical field is especially challenging since it deals with the dynamic nature of this discipline: the complexity of the human body of patients treated in modern healthcare settings. Moreover, developing and implementing effective predictive models requires a profound understanding of the data being used, along with adequate resources to properly support model development and implementation (Erdemir et al. 2020, p. 18).

Despite these challenges, predictive models have a wide range of applications in novel medical research. For example, recent studies analyzing anatomical structures via computer science have successfully used deep learning to detect vulnerable atherosclerotic plaque in preoperative patients (Toma 2023, p. 590). Similarly, computational modeling has established associations between tumor mutational burden and immunotherapy response across various cancer types (Cilla 2012, p. 59). Predictive algorithms have also been effective in identifying data patterns and forecasting surgical risks in adult spine surgery before the operation (Osorio 2016, p. 333).

Moreover, biological experiments based on computer simulation software are now referred to as "*in silico*" studies, a term derived from the silicon in computer chips, similar to the Latin terms "*in vitro*," and "*in vivo*." These novel applications of mathematical and computational models have significantly impacted the medical research field, not only in

terms of their utility and potential but also in their ability to reshape the epistemological foundations of medicine.

These new modeling environments are ideal for developing predictive individualized healthcare solutions that can result in better patient outcomes, improved patient safety, and increased drug efficacy. This concept is central to the notion of *in silico* medicine.

In recent years, computational modeling and Artificial Intelligence (AI)/Machine Learning (ML) algorithms have already been widely used in disease modeling, target identification, *in silico* trial simulations, virtual or synthetic patients, virtual coaches, and personalized medicine. While these approaches form an essential part of creating virtual replicas of physical entities, they do not encompass the entire concept of the Digital Twin.

## 5. The Digital Twin

Digital Twins (DTs) are precise digital replicas of physical entities or systems, enabling real-time monitoring, simulation, and optimization. Originally employed by NASA during the Apollo missions in the 1960s to simulate and troubleshoot space systems remotely, DTs have since expanded into various fields, with healthcare emerging as a particularly promising area.

The actual term “digital twin” was coined in 2005 by Michael Grieves in product lifecycle management, and around 2010 NASA and John Vickers utilized the Digital Twin as a virtual model of a physical system. In the original description, a DT is characterized by three components: physical, virtual, and a connection where the virtual system is mapped to the physical system by exchanging information through a real-time data connection (Katsoulakis 2024).

The Digital Twin Consortium (2022) defined the DT as a virtual representation of real-world entities and processes synchronized at a specified frequency and fidelity. Their core characteristics are:

- DT systems transform business by accelerating holistic understanding, optimal decision-making, and effective action.
- DTs use real-time and historical data to represent the past and present and simulate predicted futures.
- DTs are motivated by outcomes, tailored to use cases, powered by integration, built on data, guided by domain knowledge, and



implemented in Information Technology (IT)/Operational Technology (OT) systems.

The foundational elements of the definition are captured in the first sentence: the virtual representation, the real-world entities and processes it represents, and the mechanism by which the virtual and real-world entities are synchronized.

These definitions are consistent with the model notion explained above since they are representations of a targeted system, they can be used to make surrogate inferences, and they use mathematical and computational methods.

## 6. Digital Twins in Healthcare

In healthcare, DTs, known as Digital Twins for Health (DT4H), are transformative tools that can replicate anything from individual organs to entire hospital systems, enabling better diagnosis, treatment, and healthcare management. The development of DT4H is driven by advances in data science, AI, and the Internet of Things (IoT), which allow the creation of detailed, dynamic, and real-time models of biological systems.

Katsoulakis et al. (2024) envision that DTs should be individualized, interconnected, interactive, informative, and impactful (5Is). In the context of a patient-specific DT, the simulation, prediction, and analysis generated by the DT can be used to achieve better treatment outcomes with fewer adverse effects. Conversely, real-world data from the patient can be used to benchmark, validate, and improve the DT model.

Wakefield (2022) defines DTs as exact replicas of physical entities designed to improve or provide feedback to their real-life counterparts, similar to surrogate reasoning in other scientific models. He highlights the value of DTs in healthcare, particularly within the domain of *in silico* medicine.

Continuous improvements in product design and engineering activities have made it possible to build personalized models for patients, which can be continuously adjusted based on tracked health and lifestyle parameters. This approach can lead to the development of a virtual patient with a detailed description of the individual's healthy state, not only based on previous records but also by comparing statistics to the population to more easily identify patterns.

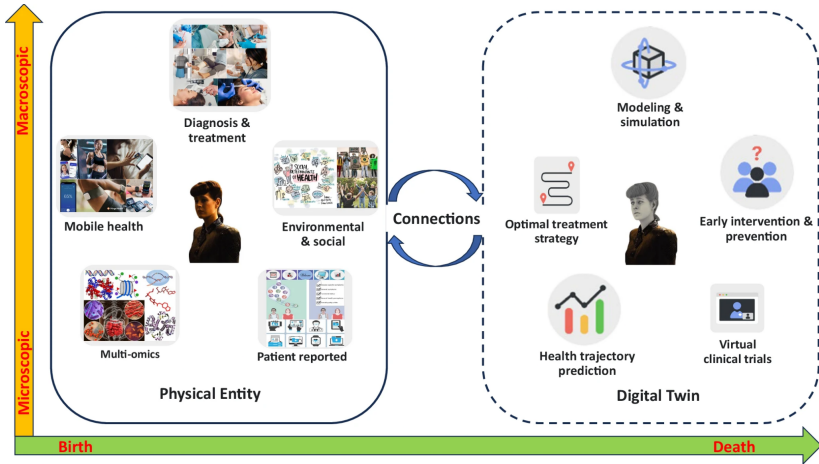


Figure 1: Digital Twin for health (DT4H) envisioned. This general scheme overviews the key factors that this model should comprehend. Both the physical entity and the Digital Twin have their intrinsic characteristics and are set in a specific scale and time (Katsoulakis 2024, p. 2).

The applications of DTs in medicine are diverse, ranging from representations of tissues, organs, organisms, patients, communities, and even healthcare systems. For example, DTs are used in hospital management and care coordination to create digital representations of healthcare data, hospital environments, operational staff, and lab results, optimizing resource utilization, improving workflow efficiency, and enhancing patient care.

In medical device design, DTs help customize devices such as those developed in the SIMULIA Living Heart project, which uses a DT of the human heart to study drug interactions and develop cardiac devices like pacemakers (Baillargeon 2014, p. 48).

In surgical planning, DTs allow surgeons to simulate procedures on virtual models before actual surgery (Sun 2022, p. 9), which is particularly useful in complex cardiac procedures, orthopedics, and cancer treatments. In drug discovery and biomarker identification, DTs accelerate the process by simulating biochemical reactions and predicting drug efficacy, reducing the time and cost involved in bringing new drugs to market (Rifaioğlu 2019, p. 20).

DTs are also used to simulate clinical trials virtually, allowing researchers to test the efficacy of treatments on virtual patients. This approach can reduce the cost and time associated with traditional clinical trials and addresses the issues concerning the recruitment of patients in the trials, which is particularly useful in fields like oncology (Kolla 2021).

Furthermore, DTs enable personalized medicine, integrating clinical history and genomic profiling by simulating the effects of different therapies on individual patients, predicting responses to chemotherapy and other treatments (Katosulakis et al. 2020).

Also, DTs have been employed to monitor and promote general health and well-being. For example, they can track fitness data from wearable devices and provide personalized health recommendations to improve overall wellness (Sahal et al. 2022, p. 22).

Finally, the Virtual Physiological Human (VPH) initiative, through the DISCIPULUS project of the European Commission, over a decade of development, concluded in the Digital Twin roadmap. They envision an ideal of a patient’s medical data, which enables the database of the VPH to be transformed into the avatar of a specific patient. Modeling studies of the avatar are guided by population-level data in the medical literature (DISCIPULUS Project 2012).

## 7. Discussion

The use of models in medical science has already proven to be useful and still remains a promising field. It not only encompasses a novel research methodology but also aligns with the modelistic era of scientific progress and has the potential to change the current paradigm of healthcare.

Notwithstanding, the development and application of the DTs arise challenges concerning its introduction into mainstream science. From a methodological and an epistemological perspective, we recognize some of the main challenges.

- **What kind of representation is involved in DTs?** As a new paradigm in modeling, it is philosophically unclear how conclusions can be drawn from mathematical models and how this information is extrapolated to the complex and dynamic nature of human physiology. The strength of the inference itself poses a logical challenge.

- **How secure is the information used by the DTs?** They rely heavily on vast amounts of personal health data to create accurate models. This raises concerns about patient privacy and the potential misuse of sensitive medical information. Securing this data from cyberattacks and unauthorized access is critical but challenging.
- **Can we rely on what DTs mean?** The accuracy of a DT depends on the quality and completeness of the data used to create it. Incomplete, outdated, or inaccurate data could lead to erroneous conclusions or predictions, potentially harming patients by suggesting incorrect treatments or interventions.
- **How ethical is the use of DTs?** The use of digital twins in healthcare raises ethical questions about consent, especially regarding how patient data is used and shared. Patients must be fully informed and give explicit consent for the use of their data in creating and maintaining digital twins. Additionally, concerns about equitable access arise as advanced digital health tools may not be available to all populations, potentially exacerbating health disparities.
- **How valid are DTs in clinical contexts?** The effectiveness and safety of using digital twins for medical decision-making must be thoroughly validated through clinical trials and evidence-based research. Without rigorous validation, the use of digital twins could introduce risks, particularly if used to replace or guide critical medical decisions.

Some of these challenges have just begun to be addressed, like the Ten Rules to create a credible practice of modeling and simulation in healthcare published by Erdermir et al. (2020): to define context clearly, to use contextually appropriate data, to evaluate within context, to list limitations explicitly, to use version control, to document appropriately, to disseminate broadly, to get independent reviews, to test competing implementations, and finally to conform to standards.

Can a new epistemological paradigm be born in medical research? Away from the solid and constant weight of medical tradition and work and in hand with the modern notion of medicine based on evidence? Only time and scientific scrutiny will tell whether this new paradigm of Medicine Based on Prediction could further develop.

## 8. Conclusion

The integration of models, particularly Digital Twins (DTs), into medical science represents a significant shift in the way healthcare is approached, researched, and practiced. Historically, models have been indispensable in simplifying and understanding complex systems, from the humoral theories of ancient medicine to the predictive algorithms of modern biomedicine. Digital Twins, however, push the boundaries of this tradition by offering dynamic, individualized simulations that have the potential to enhance decision-making, improve patient outcomes, and streamline drug discovery.

As DT4H continues to develop, it introduces not only groundbreaking opportunities but also critical challenges. From methodological concerns regarding the accuracy and representation of complex biological systems to ethical dilemmas around data security, privacy, and equitable access, the path forward is as complex as it is promising. For DTs to be successfully integrated into mainstream medicine, they must undergo rigorous validation, be supported by high-quality data, and be deployed with careful consideration of their broader social and ethical implications.

Despite these hurdles, the potential of Digital Twins to revolutionize healthcare is undeniable. Their ability to simulate patient-specific conditions, predict treatment responses, and optimize healthcare delivery could mark the beginning of a new era in medicine—one based not only on historical evidence but also on real-time prediction and personalization. The future of medical science may well rest on how effectively these models can bridge the gap between digital representation and clinical reality, shaping a new paradigm in healthcare that is more precise, individualized, and dynamic than ever before.

## Referencias

Alinier, G., Hunt, B., Gordon, R., & Harwood, C. (2006). Effectiveness of intermediate-fidelity simulation training technology in undergraduate nursing education. *Journal of Advanced Nursing*, 54(3) pp. 359-369. doi:10.1111/j.1365-2648.2006.03810.x

Baillargeon, B., Rebelo, N., Fox, D. D., Taylor, R. L., & Kuhl, E. (2014). The Living Heart Project: A robust and integrative simulator for human heart function. *Eur. J. Mech. A Solids*, 48 pp. 38-47.

Cassini, Alejandro. (2016). “Modelos científicos”. En *Diccionario Interdisciplinar Austral*, editado por Claudia E. Vanney, Ignacio Silva, y Juan F. Franck. URL=[http://dia.austral.edu.ar/Modelos\\_cientificos](http://dia.austral.edu.ar/Modelos_cientificos).

Digital Twin Consortium. (2020). Digital twin definition. <https://www.digitaltwinconsortium.org/2020/12/digital-twin-consortium-defines-digital-twin/>. (Accessed 30 July 2024)

Cilla, M., Martinez, J., Peña, E., & Martinez, M. A. (2012). Machine learning techniques as a helpful tool toward determination of plaque vulnerability. *IEEE Transactions on Biomedical Engineering*, 59(4), 1155e1161. <https://doi.org/10.1109/TBME.2012.2185495>

DISCIPULUS Project. (2012). The digital patient: Roadmap. <https://cordis.europa.eu/docs/projects/cnect/3/288143/080/deliverables/001-discipulusleafletextendedversion1.pdf>. (Accessed 30 July 2024)

Erdemir et al. *J Transl Med*, 18(369), 2020. <https://doi.org/10.1186/s12967-020-02540-4>

Frigg, Roman, and Stephan Hartmann. “Models in Science.” The Stanford Encyclopedia of Philosophy (Fall 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.). URL=<https://plato.stanford.edu/archives/fall2024/entries/models-science/>.

Hodges, Wilfrid. “Model Theory.” The Stanford Encyclopedia of Philosophy (Fall 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.). URL=<https://plato.stanford.edu/archives/fall2023/entries/model-theory/>.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4) , pp. 500-544. doi:10.1113/jphysiol.1952.sp004764

Katsoulakis, E., Wang, Q., Wu, H., et al. (2024). Digital twins for health: a scoping review. *npj Digit. Med.*, 7, p. 77. <https://doi.org/10.1038/s41746-024-01073-0>

Kermack, W. O., & McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A Containing Papers of a Mathematical and Physical Character*, 115(772) , pp. 700-721. doi:10.1098/rspa.1927.0118

Kolla, L., Gruber, F. K., Khalid, O., Hill, C., & Parikh, R. B. (2021). The case for AI-driven cancer clinical trials - The efficacy arm *in silico*. *Biochim Biophys. Acta Rev. Cancer*, 1876, 188572.

Lamas, A. (2016). *Historia de la Medicina: Hechos y personajes* (2da Edición). Editorial Mediterráneo, Chile, p. 82-83.

Michaelis, L., & Menten, M. L. (1913). Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift*, 49 pp. 333-369.

Osorio, J.A., Scheer, J.K., & Ames, C.P. (2016). Predictive modeling of complications. *Curr. Rev. Musculoskelet. Med.*, 9 pp. 333–337.

Paul, G. (2023). *Digital Human Modeling and Medicine: The Digital Twin*. ACADEMIC PRESS, Elsevier International.

Redmond, J. (2022). El desafío de razonar sustitutivamente en la práctica de modelización en ciencia. *Cuadernos Filosóficos*, 19.

Rifaioglu, A. S., et al. (2019). Recent applications of deep learning and machine intelligence on *in silico* drug discovery: methods, tools, and databases. *Brief. Bioinform*, 20, pp. 1878–1912.

Sahal, R., Alsamhi, S. H., & Brown, K. N. (2022). Personal Digital Twin: A Close Look into the Present and a Step towards the Future of Personalized Healthcare Industry. *Sensors (Basel)*, 22. <https://doi.org/10.3390/s22155918>.

Simeoni, M., Magni, P., Cammia, C., De Nicolao, G., Croci, V., Pessenti, E., Germani, M., Poggesi, I., & Rocchetti, M. (2004). Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Research*, 64(3) , pp. 1094-1101. doi:10.1158/0008-5472.CAN-03-2524

Sun, T., He, X., Song, X., Shu, L., & Li, Z. (2022). The Digital Twin in Medicine: A Key to the Future of Healthcare? *FrontMed (Lausanne)*, 9, 907066.

Toma, M., Wei, O.C. (2023). Predictive Modeling in Medicine. *Encyclopedia*, 3 , pp. 590–601. <https://doi.org/10.3390/encyclopedia3020042>

Wakefield, J. (2022). Why you may have a thinking digital twin within a decade. *BBC News*. <https://www.bbc.com/news/business-61742884>.



# La proyección antropocéntrica del *Test de Turing* en la cultura de la IA

Anllely García Rojas<sup>a</sup>

## Resumen

El *Test de Turing* sigue desempeñando un papel filosófico importante (Moor 2001), especialmente en las investigaciones y reflexiones sobre la capacidad que posee una inteligencia artificial para engañar a las personas haciéndoles creer que es humano. Estas exploraciones pueden llegar a impactar significativamente en las narrativas y/o imaginarios colectivos, esto es, en la cultura científica en torno a la IA. De hecho, la posibilidad de que sistemas de IA pasen el *Test de Turing* en muchas ocasiones despierta un temor, pero también una profunda fascinación: el sueño de la conciencia artificial y la pesadilla de la rebelión de los robots ha estado presente en gran parte de la ciencia ficción, pero también en informes periodísticos y debates públicos (Natale 2021). Surge así, la pregunta que guiará el proyecto de ensayo: ¿En qué sentido el *Test de Turing* ha afectado la cultura en torno a la IA?

El presente ensayo pretende mostrar que el *Test de Turing* contiene la idea de que la IA reside no sólo en las técnicas de programación, sino también en las formas en que los humanos perciben e interactúan con el computador. El primer apartado expone la argumentación de Simone Natale (2021) en torno a la tesis de que el *Test de Turing* define la IA en términos de la perspectiva de los humanos, en sus interacciones con los computadores, otorgando un lugar destacado a la comunicación. El segundo apartado profundiza en el impacto del *Test de Turing* –y su sesgo antropocéntrico– en el contexto sociocultural, considerando también la perspectiva de Benjamin Bratton (2015), quien enuncia la necesidad de atender a la cultura de la IA o concepción popular de la IA, comprendida como la intersección entre la IA y la cultura humana.

---

<sup>a</sup>Universidad de Valparaíso, Valparaíso, Chile.  
Contacto: anllelygarciarojas@gmail.com.

Se concluye que el *Test de Turing* proporciona claves interpretativas útiles para reflexionar acerca del impacto de la IA en las interacciones humano-computador.

**Palabras clave:** Inteligencia Artificial – Sesgo antropocéntrico – Cultura humana – Interacción humano-computador – Imaginarios colectivos.

## Abstract

The Turing Test continues to play an important philosophical role (Moor 2001), especially in research and reflections on the ability of artificial intelligence to deceive people into believing it is human. These explorations can significantly impact narratives and/or collective imaginations, that is, the scientific culture surrounding AI. In fact, the possibility of AI systems passing the Turing Test often arouses fear, but also deep fascination: the dream of artificial consciousness and the nightmare of robot rebellion has been present in much of science fiction, but also in journalistic reports and public debates (Natale 2021). Thus, the question that will guide the essay project arises: In what sense has the Turing Test affected the culture surrounding AI?

This essay aims to show that the Turing Test contains the idea that AI resides not only in programming techniques, but also in the ways in which humans perceive and interact with the computer. The first section presents Simone Natale's (2021) argument around the thesis that the Turing Test defines AI in terms of the perspective of humans, in their interactions with computers, giving a prominent place to communication. The second section delves into the impact of the Turing Test –and its anthropocentric bias– in the sociocultural context, also considering the perspective of Benjamin Bratton (2015), who enunciates the need to attend to the culture of AI or popular conception of AI, understood as the intersection between AI and human culture. It is concluded that the Turing Test provides useful interpretive keys to reflect on the impact of AI on human-computer interactions.

**Keywords:** Artificial Intelligence – Anthropocentric bias – Human culture – Human-computer interaction – Collective imaginaries.

## 1. Introducción

El matemático británico Alan Turing (1912-1954) propuso el *Test de Turing* como reemplazo de la pregunta “¿Pueden pensar las máquinas?” en “Computing Machinery and Intelligence” (1950). Esto debido a la dificultad de hallar un consenso sobre el significado de términos como ‘máquina’, ‘pensar’ e ‘inteligencia’ y también a la imposibilidad de evaluar la experiencia o la conciencia de la máquina sin estar dentro o serlo. Un animado debate comenzó poco después de la publicación del artículo y continúa hoy, a través de críticas feroces, aprobaciones

entusiastas e interpretaciones contrastantes. Esto permite vislumbrar distintos hilos argumentativos que la propuesta de Turing abrió para el naciente campo de la IA, obligando a los lectores y comentaristas a considerar las posibilidades de inteligencia artificial en el ámbito formal académico y no exclusivamente desde la literatura de ciencia ficción (Natale 2021). Es relevante mencionar que las contribuciones de Turing fueron decisivas para el posterior surgimiento de la investigación en IA. Influyó directamente en las teorías de los matemáticos John McCarthy y Marvin Minsky, quienes en 1956 organizaron la conferencia “Dartmouth Research Project on Artificial Intelligence”, donde se llegó a un consenso respecto al nombramiento e inauguración de la disciplina.

Actualmente, el *Test de Turing* sigue desempeñando un papel filosófico importante (Moor 2001), especialmente en las investigaciones y reflexiones sobre la capacidad que posee la inteligencia artificial para engañar a las personas haciéndoles creer que es humana. Estas exploraciones pueden llegar a impactar significativamente en las narrativas y/o imaginarios colectivos, esto es, en la cultura o concepción popular de la IA. De hecho, la posibilidad de que sistemas de IA pasen el *Test de Turing* en muchas ocasiones despierta temor, pero también profunda fascinación: el sueño de la conciencia artificial y la pesadilla de la rebelión de los robots ha estado presente en gran parte de la ciencia ficción, pero también en informes periodísticos y debates públicos (Natale 2021). Surge así, la pregunta que guiará el proyecto de ensayo: ¿En qué sentido el *Test de Turing* ha afectado la cultura en torno a la IA?

El presente ensayo pretende mostrar que el *Test de Turing* contiene la idea de que la IA reside no sólo en las técnicas de programación, sino también en las formas en que los humanos perciben e interactúan con el computador. El primer apartado expone la argumentación de Simone Natale (2021) en torno a la tesis de que el *Test de Turing* define la IA en términos de la perspectiva de los humanos, en sus interacciones con los computadores, otorgando un lugar destacado a la comunicación. El segundo apartado profundiza en el impacto del *Test de Turing* –y su sesgo antropocéntrico– en el contexto sociocultural, considerando también la perspectiva de Benjamin Bratton (2015), quien enuncia la necesidad de atender a la cultura de la IA o concepción popular de la IA, comprendida como la intersección entre la IA y la cultura humana. Se concluye que el *Test de Turing* proporciona claves interpretativas útiles para reflexionar acerca del impacto de la IA en las interacciones humano-computador.

## 2. La dimensión comunicacional del *Test de Turing*

En 1950, Alan Turing publicó su famoso artículo “Computing Machinery and Intelligence”, en la revista *Mind*. Una de las contribuciones más significativas para las investigaciones de las principales ramificaciones de la Ingeniería Informática; la teoría de la informática, el diseño de hardware-software y el estudio de la inteligencia artificial (Shieber 2004). En la propuesta teórica, Turing describió lo que actualmente llamamos *Test de Turing*, y al que se refirió mediante el Juego de la Imitación, del cual existen diversas versiones. No obstante, el presente proyecto refiere a la interpretación estándar, comprendida como un método empírico que evalúa si una máquina –un computador digital– puede exhibir un comportamiento inteligente.

El *Test de Turing* involucra a tres jugadores que participan en actos de comunicación: el computador digital, un jugador humano y un interrogador o juez humano. Este último tiene el desafío de distinguir cuál de los otros dos jugadores es la persona y cuál es la máquina programada a través de mensajes escritos –Turing imaginó llevar a cabo estas conversaciones por medio de una teleimpresora– (Copeland et al 2017). Los tres participantes se mantienen estrictamente fuera de contacto entre sí, sin espiar. El interrogador hace preguntas tan penetrantes y amplias como guste, y al computador digital se le permite hacer todo lo posible para forzar una identificación incorrecta. La máquina usará “todo tipo de trucos para parecerse más a un hombre”, expresó Turing en la discusión que mantuvo con Newman, R.B Braithwaite y G. Jefferson, la cual fue grabada y transmitida por BBC radio en 1952. Turing precisa respecto a estas temáticas de la siguiente forma:

BRAITHWAITE: Would the questions have to be sums, or could I ask it what it had had for breakfast?

TURING: Oh yes, anything. And the questions don't really have to be questions, any more than questions in a law court are really questions. You know the sort of thing. “I put it to you that you are only pretending to be a man” would be quite in order. Likewise the machine would be permitted all sorts of tricks so as to appear more man-like, such as waiting a bit before giving the answer, or making spelling mistakes, but it can't make smudges on the paper, any more than one can send smudges by telegraph. We had better suppose that each jury has to judge quite a number of times, and that sometimes they really are dealing with a man and not

a machine. That will prevent them saying “It must be a machine” every time without proper consideration. (Shieber 2004, pp. 118-119)

A continuación, se expone el ejemplo de Turing del tipo de conversación que podría ocurrir entre un juez y una computadora que evade con éxito la identificación:

JUDGE: In the first line of your sonnet which reads “Shall I compare thee to a summer’s day”, would not “a spring day” do as well or better?

MACHINE: It wouldn’t scan.

JUDGE: How about “a winter’s day”? That would scan all right.

MACHINE: Yes, but nobody wants to be compared to a winter’s day.

JUDGE: Would you say Mr Pickwick reminded you of Christmas?

MACHINE: In a way.

JUDGE: Yet Christmas is a winter’s day, and I do not think Mr Pickwick would mind the comparison.

MACHINE: I don’t think you’re serious. By a winter’s day one means a typical winter’s day, rather than a special one like Christmas. (Copeland et al 2017, p. 270)

El propósito del computador digital es engañar al interrogador haciéndole creer que es un humano mediante la imitación del comportamiento verbal. Mientras que el otro participante busca ayudar al interrogador revelando su identidad humana. Si el interrogador no logra hacer la correcta identificación entre el jugador humano y el computador digital, entonces este último supera la prueba. De este modo, el objetivo del *Test de Turing* es verificar si existe inteligencia mediante la imitación eficaz (González 2007). Tras definir el Juego de la Imitación, la nueva agenda a discutir se trasladó de “¿Pueden pensar las máquinas?” a la pregunta “¿Existen computadoras digitales imaginables que tendrían un buen desempeño en el juego de la imitación?” (Turing 1950, p. 442).

El artículo de Turing fue ambiguo en muchos aspectos, lo cual favoreció el surgimiento de diferentes puntos de vista y controversias sobre

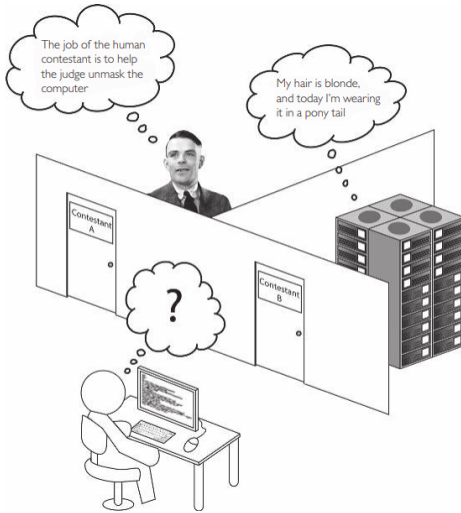


Figura 2: El *Test de Turing*. El juez debe decidir qué concursante es el computador (Copeland et al 2017).

el significado del *Test de Turing*. Sin embargo, una de las implicaciones clave para el campo de la IA es evidente. La pregunta, dice Turing a sus lectores, no es si las máquinas son o no capaces de pensar. Es, en cambio, si creemos que las máquinas son capaces de pensar, en otras palabras, si estamos preparados para aceptar el comportamiento de las máquinas como inteligente (Natale 2021, p. 20).

Simone Natale, doctor en estudios de Comunicación y Medios, sostiene en *Deceitful Media: Artificial Intelligence and Social Life After the Turing Test* que la capacidad del computador para imitar la forma en que se comportan los humanos en una conversación es obviamente uno de los factores para superar el *Test de Turing*, pero también lo es la participación activa de los jugadores humanos. Tanto el criterio utilizado por el interrogador o juez como las motivaciones personales del jugador humano que interviene en la conversación son variables que pueden influir el resultado de la prueba (Natale 2021, pp. 20-21).

Things such as their backgrounds, biases, characters, genders, and political opinions will have a role in both the decisions of the interrogator and the behavior of the human

agents with which the interrogator interacts. A computer scientist with knowledge and experience on AI, for instance, will be in a different position from someone who has limited insight into the topic. Likewise, human players who act as conversation partners in the test will have their own motivations and ideas on how to participate in the test. Some, for instance, could be fascinated by the possibility of being exchanged for a computer and therefore tempted to create ambiguity about their identities. Because of the role human actors play in the Turing test, all these are variables that may inform its outcome (Natale 2021, pp. 20-21).

Jennifer Rhee señala perspicazmente en “La promesa de la identificación errónea” que en el *Test de Turing* la responsabilidad del éxito o falla no recae únicamente en las habilidades de la máquina, sino que en gran medida recae en el ser humano (Rhee 2010). A partir de esto, es posible afirmar que los elementos decisivos para el resultado de la prueba comprenden un sesgo antropocéntrico, arraigado en la comprensión de la noción de inteligencia, más bien en la expectativa y percepción de un comportamiento inteligente similar al humano. En este sentido, la decisión de Turing respecto al diseño de la prueba, particularmente del papel del interrogador humano, se basa en un sesgo antropocéntrico. Este sesgo humano se evidencia al considerar que “(...) Turing no considera el reemplazo de los interrogadores, o que haya un programa que desempeñe la misma función que los mismos” (González 2016).

Para ilustrar la magnitud del impacto que tiene este sesgo en el resultado del *Test de Turing*, se presenta el caso de los resultados del “Premio Loebner” en el año 2000, un concurso anual en que diversos programas de computadores se involucran en una versión de la prueba:

In the 2000 contest, for example, members of the jury judged a human as a machine ten times, but did not judge any machine as a human being; and in the 2003 contest they judged a human as “definitely a machine” four times, but did not judge any computer as “definitely a human” (Copeland et al 2017, p. 290)

Al enfocarse no en el ser humano, no en la máquina, sino en el umbral y/o nexos entre estos dos agentes, es posible interpretar nuevas significaciones relacionales, no a través de definiciones estáticas e independientes, sino más bien a través de las interacciones y encuentros

entre humanos y máquinas (Rhee 2010). La lectura de Natale sostiene que, al definir la inteligencia de una máquina en términos de la capacidad de un computador para superar la prueba, Turing incluyó a los humanos en la fórmula, haciendo que sus ideas y sesgos, así como su psicología y carácter, sean una variable crucial en la construcción de máquinas “inteligentes”.

En otras palabras, independientemente de las intenciones iniciales de Turing, el *Test de Turing* ofrece definir la inteligencia artificial en términos de la perspectiva humana en su interacción con los computadores digitales. Al ser así, la centralidad de los seres humanos sugiere que la inteligencia artificial reside no sólo en circuitos y técnicas de programación, sino también en las formas en que los humanos perciben e interactúan con las máquinas. Esta interpretación, según Natale, estimuló a investigadores y desarrolladores de tecnologías a modelar comportamientos humanos y estados de ánimo con el fin de diseñar sistemas de IA interactivos más efectivos (Natale 2021).

The uncertainty that derives from this is often indicated to be one of the test’s shortcomings. The test appears entirely logical and justified, however, if one sees it not as an evaluation of the existence of thinking machines but as a measure of humans’ reactions to communications with machines that exhibit intelligent behavior. From this point of view, the test made clear for the first time, decades before the emergence of online communities, social media bots, and voice assistants, that AI not only is a matter of computer power and programming techniques but also resides—and perhaps especially—in the perceptions and patterns of interaction through which humans engage with computers (Natale 2021, p. 21).

Al situar a los humanos en el centro del diseño del *Test de Turing*, el teórico británico proporcionó un complejo escenario que surge de las interacciones humano-computador, situando a la comunicación—y una forma particular de interacción social engañosa—como un factor decisivo. Así, las tecnologías de IA contemporáneas podrían ser concebidas en términos de su credibilidad para los usuarios humanos, desde su dimensión relacional que emerge especialmente dentro de la interacción entre humanos y máquinas. Brian Christian, en un reflexivo libro en torno a su experiencia como participante en el “Premio Loebner”, destaca el aspecto comunicacional del *Test de Turing*:



Here's the thing: beyond its use as a technological benchmark, beyond even the philosophical, biological, and moral questions it poses, the Turing test is, at bottom, about the act of communication. I see its deepest questions as practical ones: How do we connect meaningfully with each other, as meaningfully as possible, within the limits of language and time? How does empathy work? What is the process by which someone comes into our life and comes to mean something to us? These, to me, are the test's most central questions—the most central questions of being human (Christian 2011, p. 9).

Natale sostiene que el diseño del *Test de Turing* se basa en las interacciones entre humanos y computadores, destacando que Turing experimentó la urgencia de incluir detalles precisos acerca del medio de comunicación; justamente con la finalidad de garantizar la validez de la prueba, el interrogador requiere comunicarse con el jugador humano y el computador sin recibir ninguna pista sobre sus identidades, excepto el contenido de los mensajes escritos. En este sentido, la comunicación entre los humanos y el computador en el *Test de Turing* estaba destinada a tener un carácter anónimo e incorpóreo (Natale 2021). En palabras de Turing:

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary (Turing 1950, p. 434).

La mediación tecnológica del *Test de Turing*, a saber, la teleimpresora, aspira a lograr que tanto los participantes digitales como los humanos participen en la comunicación como información pura, es decir, como un conjunto de datos independiente del autor. Esto sugiere que la solución dada por Turing implica una comprensión aguda del papel de los medios de comunicación. En la actualidad, el uso de diferentes tecnologías e interfaces para enviar mensajes evidencia tanto la naturaleza computacional de esos programas como el efecto de su comunicación, es decir, cómo impactan en los usuarios humanos (Natale 2021). Aquí un ejemplo, utilizado por Natale, que realiza el impacto de los medios e interfaces en los efectos de la comunicación:

Similarly, where voice is concerned, the nature of communication will be different if AI voice assistants are applied to domestic environments, like Alexa, embedded in smartphones and other mobile devices, like Siri, or employed to conduct telephonic conversations, as in automated customer services or Google's in-progress Duplex project (Natale 2021, p. 24).

Desde la perspectiva de Natale, el *Test de Turing* no solo propone definir la inteligencia artificial en términos de la perspectiva humana, esto es, desde su dimensión relacional, sino también puede ser comprendida como medio que posibilita y regula dinámicas específicas de comunicación entre humanos y computadores. Esto es de especial importancia, ya que la computación no siempre se ha visto vinculada con la comunicación.

En 1950, cuando Turing investigaba, las formas de interacción entre los usuarios humanos y las computadoras se discutían mínimamente en tanto los computadores eran considerados principalmente como herramientas de cálculo. Así, la propuesta de Turing implicó un salto visionario en la historia de la computación y la IA que contempló un cambio significativo en las actitudes culturales (Natale 2021). Y es que el matemático británico no solo comentó sobre el desarrollo de tecnologías más funcionales, sino que también estimó un cambio cultural al expresar:

The original question, "Can machines think?" I believe to be too meaningless to deserve discussion. Nevertheless I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted. I believe further that no useful purpose is served by concealing these beliefs (Turing 1950, p. 442).

Turing tenía razón al darse cuenta de que las actitudes culturales cambiarían, como consecuencia tanto de la evolución de la informática como de las experiencias de las personas con las tecnologías. De esta manera, el *Test de Turing* define la IA en términos de la perspectiva de los humanos, en sus interacciones con los computadores, otorgando un lugar destacado a la comunicación.

### 3. El impacto del *Test de Turing* en la cultura de la IA

El *Test de Turing* estimuló animadas discusiones teóricas en áreas como la psicología, la informática y la filosofía de la mente y, por tanto, resultó ser una excelente provocación para reconocer el potencial de la informática. Como consecuencia, a principios de la década de 1990, Hugh Loebner y Robert Epstein decidieron organizar un concurso público real basado en la propuesta de Turing, al que denominaron *Premio Loebner*. Desde la perspectiva de Natale, los factores que impulsaron dicho concurso no fueron exclusivamente técnicos, sino también culturales; el desarrollo de tecnologías digitales y el Internet habían creado condiciones ideales para llevar a cabo el *Premio Loebner*, por lo que, es ante todo una cuestión de comunicación y entretenimiento; su intención principal es atraer al público. Al ser así, se puede esperar que este evento y, en general, los concursos humanos versus máquinas tendrían un impacto positivo en la cultura de la IA o concepción popular de la IA, nociones que refieren a las formas en que la IA influye en la cultura humana y viceversa.

Natale considera que la popularidad de tales eventos atestigua un temor, pero también una profunda fascinación en las culturas contemporáneas, la cual es evidente en las representaciones ficticias de la IA. El sueño de la conciencia artificial y la pesadilla de la rebelión de los robots ha estado presente en gran parte de la ciencia ficción, pero también en informes periodísticos, debates públicos y/o en actuaciones públicas, donde la IA se celebra como una forma de espectáculo, como el concurso del *Premio Loebner* (Natale 2021). La difusión de estos imaginarios puede aumentar la posibilidad de que sistemas informáticos relativamente simples se perciban como inteligentes, exagerando los logros del campo de la IA. Respecto al lugar del engaño en la interacción humano-computador y las disposiciones anímicas que incentiva, Natale expresa:

People seem to have a somehow ambivalent, yet deep, desire to see the myth of the thinking machine come true, which shapes many journalistic reports and public debates about the Turing test (...) the Loebner Prize competition says more about how people are deceived than about the current state of computing and AI. The Loebner Prize competition has become an arena of deception that places humans' liability for being deceived at center stage. Yet this does not make this contest any less interesting and useful for unders-

tanding the dynamics of communicative AI systems (Natale 2021, p. 104).

El *Test de Turing* sugiere evaluar la inteligencia del computador considerando su capacidad para simular la forma en que se comportan los humanos en una conversación mediante el engaño. La idea de que la capacidad de mentir es una característica definitoria de los humanos se encuentra en la base de la prueba. En este sentido, incluir la mentira y el engaño en la definición de la IA vuelve la atención a la centralidad de los humanos y, por tanto, al sesgo antropocéntrico y/o lo que Benjamin Bratton en *Outing Artificial Intelligence: Reckoning with Turing Tests* denominará “falacia antropocéntrica”; la creencia de que toda inteligencia debe ser similar o idéntica a la humana. Esto llevaría a pensar la IA exclusivamente en términos de la perspectiva humana.

Airplanes do not fly like birds fly, and we certainly do not try to trick birds into thinking that airplanes are birds in order to test whether those planes “really” are flying machines. Why do it for AI then? Today the vast majority of core AI research is not focusing Turing Test as anything like a central criterion of success, and yet in our general discourse about AI, the test’s anthropocentrism still holds such conceptual importance (Bratton 2015, p. 74).

La centralidad de la perspectiva humana en la definición de la IA puede ser interpretada como una expresión del antropocentrismo, el cual tiene consecuencias directas en las dinámicas humano-computador y, por consiguiente, en un contexto social y cultural. El *Test de Turing* estableció el engaño como un efecto probable de dichas interacciones, ya que el computador “ganará” el juego si es capaz de engañar al interrogador humano, es decir, hacerle creer que el computador digital es humano mediante el comportamiento verbal. En este sentido, puede ser interpretada como una prueba de detección de mentiras en la que no es posible que el interrogador confíe en sus compañeros de conversación, pues todo mensaje emitido por la máquina es falso (Natale 2021).

Es por este motivo, quizá, que la posibilidad de que los sistemas de IA superen el *Test de Turing* a menudo despierta temores; ya que las máquinas engañarán a los humanos con el fin de avanzar un paso más en su rebelión contra sus creadores (Natale 2021). De este modo, el sesgo o falacia antropocéntrica, presente en el *Test de Turing*, se proyecta en la cultura o concepción popular de la IA, reflejándose en innumerables

narrativas o ficciones que otorgan características humanas, como la dimensión afectiva o la intencionalidad –ira, celos, confusión, avaricia, orgullo, deseo, entre otros– a los sistemas de IA (Bratton 2015). Esto se encuentra dramatizado en innumerables narrativas de ciencia ficción, desde “Blade Runner” de Ridley Scott hasta “Her” de Spike Jonze.

La falacia antropocéntrica configura un prisma a través del cual gran parte de la cultura humana comprende las interacciones con la IA, por lo que Bratton recomienda un análisis crítico de sus implicaciones a nivel social, cultural y político. Debido a que, si se busca la definición de IA de manera equivocada, esta puede surgir en formas que son innecesariamente difíciles de reconocer, amplificando sus riesgos y la opacidad o mistificación del funcionamiento de la computación y, a su vez, de los logros del campo de la IA (Bratton 2015).

We need a popular culture of A.I. that is less parochial and narcissistic, one that is based on more than simply looking for a machine version of our own reflection. As a basis for staging encounters between various A.I.s and humans, that would be a deeply flawed precondition for communication. Needless to say, our historical track record with “first contacts,” even among ourselves, does not provide clear comfort that we are well-prepared (Bratton 2015).

La relación entre la IA y el engaño configura una problemática relevante para las interacciones entre humanos y computadores que suscita profundas reflexiones. Por este motivo, la preocupación de Natale se centra en el lugar del engaño, defendiendo que “la IA debería situarse dentro de la trayectoria más larga de los medios engañosos que incorporan engaño banal en su función” (Natale 2021, p. 31). El *Test de Turing* contribuye a la justificación de dicha afirmación al momento de considerar el componente lúdico del engaño administrado por el computador.

Se suele atribuir una connotación negativa al engaño, pero su integración en actividades lúdicas es un recordatorio de que las personas buscan activamente situaciones en las que pueden ser engañadas, siguiendo un deseo o necesidad compartida (Natale 2021). En este sentido, el *Test de Turing* presenta la posibilidad de que los humanos puedan ser engañados por las computadoras en un contexto lúdico y voluntario. A continuación, se expone el ejemplo de Natale que muestra que cierto grado de engaño es inofensivo e incluso funcional en una interacción cotidiana:

Alexa and Siri are perfect examples of how this works in practice: the use of human voices and names with which these “assistants” can be summoned and the consistency of their behaviors stimulate users to assign a certain personality to them. This, in turn, helps users to introduce these systems more easily into their everyday lives and domestic spaces, making them less threatening and more familiar. Voice assistants function most effectively when a form of playful and willing deception is embedded in the interaction (Natale 2021, pp. 29-30).

La proyección de la falacia antropocéntrica, materializada en el mito tecnológico que se centra exclusivamente en la posibilidad de crear máquinas pensantes en un nivel técnico, aleja la atención del hecho de que las implicaciones más significativas de los sistemas de IA no se encuentran necesariamente en un futuro lejano, sino que en las interacciones cotidianas. Esto es de suma importancia si consideramos que, en su día a día, los humanos deciden construir un espacio de interacción recíproca con diversos sistemas de IA que modifican el mundo que los circunda. Por este motivo, Natale enuncia la urgencia de reflexionar sobre el lugar del engaño interdisciplinariamente; en la encrucijada entre las ciencias de la computación, las ciencias económicas, las ciencias cognitivas, las ciencias sociales y las humanidades.

## 4. Conclusión

Un animado debate comenzó poco después de la publicación del artículo “Computing Machinery and Intelligence” (1950), con fuertes críticas, apoyo entusiasta e interpretaciones contrastantes, y continúa hasta el día de hoy. Como se demostró, el *Test de Turing* sigue desempeñando un papel filosófico importante, especialmente en las reflexiones sobre la dimensión relacional de la IA, la cual emerge especialmente dentro de la interacción humano-computador. De esta manera, aun si el *Test de Turing* no refleja el funcionamiento de los sistemas modernos de la IA, proporciona claves interpretativas útiles para comprender las comunicaciones entre humanos y máquinas, situadas en un contexto social y cultural. La influencia de la inteligencia artificial en los comportamientos sociales evidencia la relevancia de la temática desarrollada en este ensayo.

Siguiendo las reflexiones de Benjamin H. Bratton (2015) y Simone Natale (2021), la argumentación desplegada ante la pregunta: ¿En qué

sentido el *Test de Turing* ha afectado la cultura en torno a la IA? Se centró en mostrar, en el primer apartado, que independientemente de las intenciones iniciales de Turing, el *Test de Turing* contiene la idea de que la IA reside no sólo en las técnicas de programación, sino también en las formas en que los humanos perciben e interactúan con el computador. Así, la definición de inteligencia artificial depende de la percepción humana más que de alguna característica específica de las máquinas. De este modo, el *Test de Turing* no solo reconoce la dimensión relacional de la IA, sino también permite interpretarla como medio que posibilita y regula dinámicas específicas de comunicación entre humanos y computadores. El hecho de que el *Test de Turing* planteó una situación en la que un interrogador humano era propenso al engaño por la máquina suscitó una profunda controversia en el campo de la IA desde sus inicios. Por esto, en el segundo apartado se profundizó en el impacto del *Test de Turing*, particularmente desde la falacia antropocéntrica, en el contexto sociocultural, exponiendo que también se debe considerar el hecho de que la prueba se enmarca en lo lúdico y lo inofensivo de este engaño. De esta forma, aquello que se consideraba un simple objeto de la ciencia parece convertirse en un verdadero sujeto, ya que presenta la capacidad de tejer una red de relaciones con los humanos. Con todo, el presente escrito vislumbra en el *Test de Turing* un espacio de intersección entre la IA y la cultura humana, brindando un punto común para el desarrollo tecnológico de sistemas IA y los medios de comunicación o interfaces que posibilitan una reflexión más amplia de las implicaciones epistemológicas de la interacción humano-computador. Es necesario ahondar en estas temáticas desde la interdisciplinariedad, para enfrentar los desafíos éticos y políticos que surgen en este contexto.

## Referencias

- Bernhardt, C. (2016). *Turing's vision: the birth of computer science*. Cambridge, Massachusetts: The MIT Press.
- Bratton, B. (2015). Outing Artificial Intelligence: Reckoning with Turing Tests, en Pasquinelli, M. *Alleys of Your Mind. Augmented Intelligence and Its Traumas*. Lüneburg: meson press , pp. 69–80.
- Bratton, B. (2015, febrero 23). Outing A.I.: Beyond the Turing Test. *The New York Times*. Recuperado de <https://>

//archive.nytimes.com/opinionator.blogs.nytimes.com/2015/02/23/outing-a-i-beyond-the-turing-test/.

Copeland, J. (2012). *Turing: Pioneer of the Information Age*. United Kingdom: Oxford University Press.

Copeland, J., Bowen, J., Sprevak, M., & Wilson, R. (2017). *The Turing Guide*. New York: Oxford University Press.

Copeland, J. (Ed.). (2004). *The essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secret of Enigma*. Oxford: Oxford University Press.

Christian, B. (2011). *The Most Human Human: What Talking with Computers Teaches Us About What It Means to Be Alive*. New York: Doubleday.

González, R. (2007). El Test de Turing: dos mitos, un dogma. *Revista de Filosofía*, 63 pp. 37–53.

González, R. (2016). Dos criterios para la presencia de estados mentales: Descartes y Turing. *Cinta Moebio*, 56 , pp. 159–171.

Gunkel, D. (2018). *Gaming The System: Deconstructing Video Games, Games Studies, and Virtual Worlds*. Bloomington: Indiana University Press.

Moor, J. (Ed.). (2003). *The Turing Test: The Elusive Standard of Artificial Intelligence*. Nueva York: Springer Science+Business Media.

Natale, S. (2021). *Deceitful media: artificial intelligence and social life after the Turing test*. New York: Oxford University Press.

Oppy, G., & Dowe, D. (2003). The Turing Test. *The Stanford Encyclopedia of Philosophy*. Recuperado de <https://plato.stanford.edu/archives/win2021/entries/turing-test/>.

Turing, A. (1950). Computing machinery and intelligence (Trad. Fuentes 2010). *Mind*, Vol. 59, ed. 236 pp. 433–460. doi:10.1093/mind/lix.236.433

Rhee, J. (2010). Anthropomorphic Attachments in U.S. Literature, Robotics, and Artificial Intelligence. Duke University.

Shieber, M. (Ed.) (2004). *The Turing test: verbal behavior as the hallmark of intelligence*. Cambridge, Massachusetts: The MIT Press.



# Conspirativismo digital y demarcación científica: un acercamiento desde la etnografía virtual

Manuel García Domínguez<sup>a</sup> y Pablo Verde Ortega<sup>b</sup>

## Resumen

Las teorías conspirativas han gozado de un crecimiento exponencial en la última década gracias a su viralización a través de distintas plataformas digitales como Telegram. Estas comunidades conspirativas han sido criticadas por su abandono de toda forma de cientificidad desde filósofos clásicos como Popper o Bunge. Sin embargo, un estudio detallado de las comunidades conspirativas vigentes da cuenta de que en tales comunidades la ciencia sigue conservando, en última instancia, autoridad epistémica. A la luz de un análisis histórico del problema de la demarcación desde el Círculo de Viena, en este ensayo desarrollaremos cómo las comunidades conspirativas redibujan falazmente los límites que demarcan la ciencia conforme a criterios no epistémicos, sino políticos, así como emplean la ciencia terminológicamente para la defensa de ciertas ideologías. Posteriormente, relacionamos esta nueva demarcación con fenómenos emergentes en redes sociales como la posverdad, las ventanas conceptuales o las burbujas epistémicas, incidiendo en su vínculo con los vicios asociados a estas comunidades conspirativas. Tras ello, nos aproximamos desde una etnografía virtual a un grupo de Telegram conspirativo dedicado a la crítica de la Agenda 2030, considerada una trama oculta para el control de la población a través de “engaños científicos” como el cambio climático o

---

<sup>a</sup>Universidad Complutense de Madrid y Universidad Nacional de Educación a Distancia, Madrid, España.

Contacto: mangar21@ucm.es

<sup>b</sup>Universidad Complutense de Madrid y Consejo Superior de Investigaciones Científicas, Madrid, España.

Contacto: paverde@ucm.es

la pandemia. Finalmente, tras la crítica realizada, proponemos ciertas transformaciones de la ciencia convencional de la epistemología social y feminista para reconfigurar una imagen pública de la ciencia que evite la generación de este tipo de comunidades o, al menos, la neutralice en el imaginario social.

**Palabras clave:** Conspirativismo, demarcación, ciencia, pseudociencia, posverdad

## Abstract

Conspiracy theories have enjoyed exponential growth in the last decade thanks to their viralisation through different digital platforms such as Telegram. These conspiracy communities have been criticized by classical philosophers such as Popper and Bunge for their abandonment of any form of scientificity. However, a detailed study of current conspiracy communities reveals that in such communities science ultimately retains epistemic authority. In the light of a historical analysis of the problem of demarcation since the Vienna Circle, in this essay we will develop how conspiratorial communities fallaciously redraw the boundaries demarcating science according to political rather than epistemic criteria, as well as employing science terminologically for the defense of certain ideologies. Subsequently, we relate this new demarcation to emerging phenomena in social networks such as post-truth, conceptual windows or epistemic bubbles, highlighting their link with the vices associated with these conspiratorial communities. After that, we approach from a virtual ethnography to a conspiratorial Telegram group dedicated to the criticism of the 2030 Agenda, which considers it a hidden plot for the control of the population through 'scientific deceptions' such as climate change or pandemics. Finally, following this critique, we propose certain transformations of the conventional science of social and feminist epistemology in order to reconfigure a public image of science that avoids the generation of this type of community or, at least, neutralizes it in the social imaginary.

**Keywords:** Conspirativism, demarcation, science, pseudo-science, post-truth

## 1. Introducción

En los últimos años asistimos a un aumento de las comunidades que pivotan en torno a alguna teoría de la conspiración. Tanto es así que autores como Ramonet (2023) han rotulado nuestra época como la era del conspiracionismo, donde las comunidades conspirativas se caracterizan por “hacer referencia a un plan urdido por un grupo que mantiene ocultas sus intenciones y acciones con el fin de conseguir ventajas de orden político, económico o social” (Gallo 2019, p. 219). A esto Patán (2006) ha añadido una serie de características comunes tales

como la presentación de un mundo ordenado sin casualidades, la liberación de sus seguidores de responsabilidades, el seguimiento de lógicas maniqueas y la posesión de cierta carga milenarista. En su dimensión epistémica, Popper (1994) entenderá las teorías de la conspiración como discursos cerrados y autoafirmativos y los despoja de todo carácter de cientificidad al no poder falsar aquello que refiere a lo constantemente oculto. Ante esto cabría recuperar una postura epistemológica particularista ante las teorías de la conspiración, es decir, no considerar que estas sean por defecto descartables o epistémicamente erradas y defender que si llegan a acumular un grado considerable de evidencia a su favor, se pueden tener en cuenta como una forma legítima de explicación de ciertos fenómenos (Verde 2022). Coherentemente, un aumento de las tendencias conspirativas no debería ser un motivo intrínseco de preocupación. Sí lo es, no obstante, el hecho de que este aumento esté en su gran mayoría ligado a grupos que ponen su acento conspirativo en temas como la pandemia del SARS-COV-19 o el cambio climático, que no solo han sido y son de vital importancia sociopolítica, sino que además cuentan con un respaldo científico abrumador a sus espaldas.

Nuestro objeto en el presente artículo es centrarnos en el caso particular de los conspiracionistas englobados en torno a la crítica a Agenda 2030, analizarlo en lo que tiene de singular y cotejar cuáles son los vicios epistémicos que lo lastran (pero de los que no necesariamente adolecen las demás teorías de la conspiración) y que a la postre hacen de la misma una teoría errada, insatisfactoria e incluso perjudicial. Finalmente, ofreceremos algunas conclusiones de la mano de la epistemología social y feminista contemporánea que podrían sentar las bases de una cultura de la discusión pública de la ciencia alejada tanto del negacionismo más deletéreo como del dogmatismo y la confianza ciega en las instituciones científicas. De esta manera, dispondremos de herramientas epistémicas mucho más sólidas para solucionar problemas acuciantes como fue la pandemia apenas tres años atrás y seguirá siendo el cambio climático, entre otros.

## **2. El problema de la demarcación científica y sus derivadas**

A pesar de que una de las características más comunes del conspiracionismo político es el negacionismo científico, es decir, el rechazo de los principales consensos de la comunidad científica, no debemos creer que la ciencia está desterrada del discurso conspirativo. De hecho, la

ciencia permanece en una posición central dentro de su argumentario y aquello que distingue al conspirativo no sería tanto su renuncia a la cientificidad sino una concepción singular de la misma. Ahora bien, esta cientificidad alternativa no supone sólo una ampliación del concepto a nuevas disciplinas y estudios, sino también un cuestionamiento de ciertos criterios más asentados de cientificidad que fragmenta las ciencias establecidas y emplea los trozos resultantes a su favor. Ante esto podríamos preguntarnos ¿qué entendemos por ciencias y sobre qué sectores de tal ecosistema se vierten las críticas conspirativas?

Los límites de las ciencias se han delineado especialmente durante el último siglo a partir de los análisis del Círculo de Viena. Aunque el problema que realmente concernía a los filósofos englobados bajo este rótulo no era tanto la delimitación de la ciencia frente a la pseudociencia sino frente a la metafísica, sentaron las bases del futuro problema de la demarcación científica a partir de su teoría verificacionista: “una aseveración científica debía poder distinguirse de una metafísica por la posibilidad de ser verificada” (Hansson 2021), entendiendo por “verificación” la capacidad para ofrecer apoyo empírico a favor de una proposición dada.

Fue Karl Popper quien dio forma definitiva al problema de la demarcación en filosofía de la ciencia, que consideraba el más importante de esta disciplina. El criterio de demarcación que propuso no es otro que la falsabilidad: “para ser calificada de científica, una proposición o un conjunto de proposiciones debe ser capaz de entrar en conflicto con observaciones posibles o concebibles” (Popper 1962, p. 39). Asimismo, el austriaco fue quien desplazó el foco desde la metafísica, poseedora aún a su juicio de algún significado, a la pseudociencia. Imre Lakatos (2018), por su parte, añadió en los años 70 una capa más de profundidad al ya venerable problema de la demarcación al cambiar el foco de la teoría aislada al programa de investigación. Según el filósofo húngaro, un programa de investigación es progresivo si las nuevas teorías hacen predicciones sorprendentes que se confirman. En contraste, un programa de investigación degenerativo se caracteriza por teorías fabricadas sólo para acomodar hechos conocidos. Lakatos enfatizó que el progreso en la ciencia es posible si cada nueva teoría desarrollada en el programa tiene un contenido empírico más amplio que su predecesora.

A este respecto, Larry Laudan (1983) escribía en los años 80 sobre la muerte del problema de la demarcación:

En la medida en que nuestra preocupación es protegernos a nosotros mismos y a nuestros compañeros del pecado car-

dinal de creer en lo que deseamos que sea cierto en lugar de en lo que existe una evidencia sustancial (y seguramente eso es a lo que la mayoría de las formas de “charlatanería” se reducen), entonces nuestro enfoque debe estar completamente centrado en las credenciales empíricas y conceptuales de las afirmaciones sobre el mundo. El estado “científico” de esas afirmaciones es completamente irrelevante (p. 125).

A pesar de la perspectiva de Laudan, la filosofía de la ciencia posterior a estas afirmaciones ha continuado ofreciendo criterios de demarcación, si bien desde una perspectiva más gradualista, pragmática y sobre todo pluralista. Uno de los grandes inconvenientes en el planteamiento del problema de la demarcación por los autores clásicos (Círculo de Viena, Popper y Lakatos) residía en su insistencia en el criterio único, lo que acabó conduciendo a cada uno de los modelos demarcacionistas a aporías, contraejemplos e insuficiencias que dificultaban cualquier separación mínimamente solvente entre ciencias y pseudociencias.

Ante los callejones sin salida del monismo, encontramos a partir de los años 80 aproximaciones como la de Mario Bunge (1982), que critica los intentos previos de demarcación científica precisamente por su carácter unilateral. Al contrario, el filósofo argentino-canadiense defiende la importancia de una combinación de criterios para poder dar una respuesta sólida al secular problema de la demarcación. Así pues, ofrece en 1985 un decálogo para evaluar la (falta) de científicidad de una disciplina, del cual podemos recuperar algunos de sus puntos:

1. Es una comunidad de creyentes, no de investigadores;
2. La sociedad anfitriona apoya a la comunidad de creyentes por motivos prácticos (porque es un buen negocio o porque refuerza a la ideología imperante) o la tolera, aunque la exilia fuera de la cultura oficial;
3. El dominio o universo del discurso contiene ítems imaginarios, tales como influencias astrales, pensamientos desencarnados, super-egos, memorias ancestrales, voluntad nacional, destino manifiesto, objetos voladores no identificados y similares a los que todos los miembros de la comunidad les asignan existencia real. [...]
4. El fondo específico es muy pequeño, cuando no vacío: una seudociencia aprende poco o nada de otros campos de conocimientos y contribuye poco a nada a ellos;

5. La problemática incluye problemas mal planteados (por tener supuestos falsos) y típicamente (aunque no siempre) prácticos más que cognoscitivos.
  
6. El fondo de conocimientos acumulados es pequeño, está estancado y contiene numerosas hipótesis incontrastables o incompatibles con hipótesis científicas bien confirmadas (leyes); en particular, no contiene leyes propiamente dichas (Bunge 1985, pp. 68-69).

Los conspirativos políticos, como veremos, han contribuido a esta revitalización del problema de la demarcación aunque en una dirección opuesta al “giro pragmático” que caracteriza a la mayoría de filósofos de la ciencia contemporáneos dedicados a esta cuestión. En vez de desplegar criterios múltiples y flexibles, los conspiracionistas aspiran a alcanzar un mecanismo de triaje entre conocimientos científicos (legítimos) y pseudocientíficos (ilegítimos) que sea único y uniforme. Ya no se trata de delimitar lo científico de lo no científico, sino el conocimiento legítimo del espúreo, el que es producido por las élites globalistas frente al que ofrece resistencia a las mismas. El criterio de demarcación, como vemos, es eminentemente ideológico, político y no epistémico. Así, los conspirativistas darán por válido sólo aquel conocimiento científico que sirva para justificar sus propias teorías y praxis. Dado lo injustificable de la mayoría de ellas, estos grupúsculos deberán mutilar las ciencias existentes, trocear las diferentes teorías y seleccionar solo aquellas partes que les resulten convenientes, lo que en muchas ocasiones lleva a recontextualizar los pedazos de evidencia científica tomados de maneras que contradicen las conclusiones originales de las mismas.

He aquí los dos grandes vicios epistémicos del conspiracionismo: su recaída en las soluciones monistas al problema de la demarcación y su empleo arbitrario de los factores sociales como criterio demarcador, lo que de la mano de Bunge (2014) podríamos llamar “escepticismo radical” de corte político (p. 198). Esto no obsta para que la mayoría de conspiracionistas se vean a sí mismos como los más acérrimos defensores de la ciencia (o de lo que ellos demarcan como ciencia) hasta el punto de reproducir lo que podríamos denominar un “pseudopositivismo” que priva de validez cualquier afirmación que se haga por fuera de los márgenes que su peculiar sistema de demarcación impone.

### 3. Conspiracionismo en la era digital

Este auge del conspiracionismo tiene una estrecha relación con la transición digital de las sociedades contemporáneas, especialmente con las tecnologías de la información y la comunicación. Estas tecnologías han generado nuevos espacios para la discusión dentro del espacio público como canales semiprivados de Telegram o WhatsApp, desde donde se comparte información sin dar cuenta de sus atributos de veracidad, en lo que Dawkins vino a llamar ‘meme’. Al usuario de estas plataformas, Lara (2017) lo caracteriza por una gran incontinencia, debate huidizo, lenguaje soez, supraideas marcadas, argumentos de terceros y fuentes no contrastadas. Un contexto de sobrecarga de información supone una sobrecarga, a su vez, de información no contrastada, siendo su primera consecuencia la descentralización de las formas vigentes de verificación de la información. Esta sobrecarga de información conlleva una transformación del tiempo para el consumo y cierta confianza social en el contenido que se consume; precisamente serán esta pérdida de la autoridad previa dentro del ruido social y la centralización de la responsabilidad epistémica en un individuo temporalmente precarizado dos pilares sobre los que emerge la llamada posverdad. En palabras de Zafra (2017): “la pareja celeridad y exceso contribuye a reforzar emociones e ideas preconcebidas bajo la lógica de que no puede haber parada reflexiva sin tiempo para ello” (p. 72).

La posverdad se entiende así como aquellas circunstancias donde los “hechos objetivos son menos influyentes en la formación de la opinión pública que la apelación a las emociones y las creencias personales” (Brahms 2020, p. 1). Esta centralización del discurso emotivo sobre la información reconfigura profundamente aquello que entendemos por información científica; en este sentido no se pretende un cambio de contenido o una alteración de la forma teórica, sino que “lo que está en juego es lo que rige los enunciados y la manera en que se rigen los unos a los otros para constituir un conjunto de proposiciones aceptables científicamente” (García 2018, p. 180). Hay cierto consenso en señalar que la posverdad está estrechamente relacionada con el relativismo y que por tanto se desdibujan absolutamente los límites de la ciencia conforme a los criterios personales de cada cual. Sin embargo, esta postura plantea serias limitaciones a la hora de dar cuenta del carácter eminentemente social de las redes digitales. El relativismo parece reducir las cuestiones al individuo como individuo, mientras que en el caso del conspirativismo político el valor de verdad de las distintas cuestiones

se amplía al individuo como parte de la comunidad conspirativa con la que comparte ciertas ventanas conceptuales y que funciona como burbuja epistémica. Precisamente será esta tensión entre la formación de emociones comunes y la viralización de un exceso de información propia de ecosistemas digitales el caldo de cultivo para el desarrollo y la emergencia de la cientificidad conspirativa.

## 4. Estudio de caso

La comunidad de estudio, como veníamos tratando, es un reciente grupo hispano de Telegram organizado contra la Agenda 2030, también conocida como Agenda para el Desarrollo Sostenible, planteada por la ONU en 2015. Aunque hay cierta variedad dentro del grupo, se considera que esta agenda es un plan orquestado por una élite oculta y diabólica que pretende el control absoluto de la población a través del genocidio y la inoculación. En este sentido, en adelante se considerará esta una comunidad conspirativa al seguir las condiciones mencionadas al comienzo del ensayo: la introducción de elementos como agentes poderosos, secretos y potencialmente amenazantes, y la consideración del mundo como un lugar poco confiable y lleno de engaños que puede moldear la forma en que se seleccionan, evalúan y presentan la información (Meyer et al. 2022). Esta postura tendrá expresiones directas en la ignorancia de los métodos de producción de conocimiento científico, la simplificación de la relación entre las instituciones políticas, sociales y científicas, y la comprensión de la realidad como algo fácil de desentrañar por la experiencia personal desarraigada de sus implicaciones sociales. Veámoslo con ejemplos emitidos en el mismo grupo de Telegram.

### 4.1 La Ciencia y la ciencia para esclavos

El abordaje del hecho científico por parte del grupo de Telegram es complejo y presenta muchas aristas. Los miembros del grupo, a pesar de su heterogeneidad, se consideran abiertamente científicos, dotando de una máxima autoridad enunciativa a cierta información presentada como científica. Por ejemplo, Q., usuario del grupo, incide en que “es fácil derribar la mentira del cambio climático porque no se sostiene científicamente ni tampoco por observación”. Sin embargo, como hemos venido exponiendo, no se trata de la ciencia en su sentido usual, sino que se establece una distinción radical entre la llamada “Ciencia” y las “ciencias para esclavos”. D., usuario del grupo, dirá: “quienes nos



mienten hoy aparecen como representantes de la ciencia (ciencia para esclavos, no Ciencia) como sacerdotes, como políticos, como militares o paramilitares mandados”. La Ciencia se presenta en primer lugar bajo una definición en términos de negación, como una forma de conocimiento alejada de las instituciones tradicionales tales como el Estado o la Iglesia. En otros términos, no se dice positivamente qué es la ciencia, sino que se aspira a hacer una demarcación con base en todo aquello que no es; como comentábamos, seccionando el territorio científico y descartando aquellas teorías que no encajan con su ontología política. T., otro usuario del grupo, continúa esta crítica a la ciencia mediática:

Hace siglos, cuando la verdadera “ciencia” era un lujo de pocos, la gente hablaba de “La ira de Dios” ante estas situaciones. Hoy, con “la otra ciencia” pregonada por los medios, la ciencia real es silenciada para que los “fieles creyentes” sustituyan a un Dios por otro.

Este punto es profundamente interesante ya que se adopta desde su postura neodemarcacionista la crítica que se les plantea desde posiciones demarcacionistas clásicas. Para el mencionado García (2018) “la tecnología facilita la propagación de noticias falsas y sin fundamento y comprobación de su veracidad. Por ello se habla de una verdad científica y una posverdad mediática” (p. 187). En este sentido, desde la comunidad conspirativa se produce una desvinculación del grupo de Telegram con las fuerzas mediáticas, tratando de presentar una alternativa informativa y pudiendo plantear así esa misma crítica: aquello de que se informa en los medios de comunicación tradicionales es falso y sólo esta comunidad tiene acceso a la verdad científica. A este respecto, S., usuario del grupo, relata:

En este momento somos una fuerza social creciente en todo el mundo, por eso lo que más temen es que seamos capaces de canalizarla. Millones de personas que en su momento se inocularon ahora son de los nuestros, se han dado cuenta. Y no solo eso, se han hecho inmunes al resto de engaños colectivos como son el Canto Climático, la Guerra de Volodomir y las constantes campañas de la Afrenta 2030. La credibilidad de la ciencia oficial, de los médicos, de los periodistas está bajo mínimos, sobre todo los más jóvenes se informan en Internet. Los hacedores de todo esto saben que ya no cuela.

Se podría inferir que se considera científico cualquier información con apariencia científica que no provenga ni esté apoyada por las instituciones estatales, los principales medios de comunicación y la comunidad científica internacional. De hecho, tendrán lo que han llamado su propia comunidad científica, como dice D.: “hay una comunidad médica que lleva décadas denunciando vacunas infantiles [...] hay pruebas suficientes que las relacionan con cánceres, leucemia, parálisis...”. Ahora bien, la misma noción de “apariencia científica” es de por sí problemática y quizás haya que esbozar algunas de sus expresiones. En términos generales, se produce una banalización de la forma científica en cuanto se considera con apariencia científica todo aquello que siga las representaciones estereotipadas del científico, a saber, que sea enunciado por una persona con bata blanca que se dice científica, que incluya una sobrecarga de datos cuantitativos que parecen apoyar una hipótesis o que añada palabras complejas relacionadas con ámbitos de las ciencias naturales, tales como la biología, la física o la química (Urteaga 2013). M., usuario del grupo y autodenominado alquimista, señalará a un científico llamado Bartomeu Payeras, al que presenta erróneamente como catedrático de la Universidad de Barcelona, que investigó una oculta agenda de inoculación masiva con grafeno, lo cual en sus mismos términos “ha generado coronas bio-moleculares en las células producidas a causa de las nanopartículas y los campos CEM y produce desechos exosómicos con fragmentos de información errónea en forma de bolsitas proteicas con cadenas de ARN”. Otro ejemplo paradigmático al que citan es el físico John Clauser, autor no especializado en cuestiones climáticas, perteneciente a la C02 Coalition, organización que recibe financiación de grandes industrias petroleras. Se produce así una reducción del contenido científico a sus conclusiones, característica central de toda pseudociencia, sin dar relevancia al método o los autores encargados de la producción de conocimiento. En palabras de Bunge, “contiene métodos que no son contrastables ni justificables; en cambio, típicamente, la pseudociencia no hace experimentos ni admite la crítica” (Bunge 1985, p. 70). La diferencia entre Ciencia y ciencia para esclavos reside en que si bien ambas tienen una forma científica, la primera cumple una segunda condición: confrontar con los consensos colectivos de la comunidad científica y enfrentarse con una agenda oculta, exponiendo una suerte de verdad revelada. Hacia esta idea apuntará Popper (1945) cuando describa la quinta columna de las teorías conspirativas como la secularización de las supersticiones religiosas.

## 4.2 La cuestión del Canto Climático y la agencia política

Para aterrizar el análisis de la cuestión científica en esta comunidad conviene posarnos sobre un tema en concreto. En este caso, será interesante abordar la cuestión del cambio climático. El cambio climático es un consenso de la comunidad científica, la mayor parte de la sociedad civil y las instituciones. Si bien se dan distintas formas de negacionismo, es inusual encontrar lo que Riechmann ha venido a llamar negacionistas de nivel cero o nivel uno (Riechmann 2020), es decir, aquel negacionismo que rechaza respectivamente la mera existencia de un cambio climático o lo asocian a un cambio no antropogénico. Muchos de tales negacionistas se agrupan en este tipo de comunidades donde se considera que el cambio climático es una invención de la Agenda Globalista de Davos y la ONU que defenderían una ideología, según A., usuario del grupo, “post-izda woke-globalista”. De hecho, en relación con la Agenda 2030 es especialmente relevante ya que sitúan el cambio climático como fenómeno central en el desarrollo de la agenda. A., siguiendo esta línea, escribe:

La agenda es muy fácil derribarla ya que toda está basada realmente en una sola mentira que es la mentira del cambio climático, fíjense que todos [s.c. los puntos de la Agenda 2030] en su esencia tienen como excusa el cambio climático, así que derribando esa mentira que es muy fácil de destruir en realidad toda la satánica agenda caerá.

En este sentido, hay dos líneas principales: aquella que defiende que el cambio climático está siendo originado deliberadamente mediante lo que entienden como geoingeniería para el control de la producción agraria, y aquella que niega que los datos sean ciertos, es decir, que efectivamente este cambio no existe. Por supuesto, esto sólo puede considerarse entendiendo el cambio climático como un fenómeno aislado con nula conexión con otros fenómenos de la crisis ecosocial como la acidificación del agua oceánica, la transformación del suelo, la perturbación de los flujos bioquímicos... Ante la muerte de los grandes relatos, la forma de dar un sentido a la totalidad del mundo es la explicación de cada fenómeno bajo un microrrelato, fragmentando radicalmente una realidad materialmente interconectada. Esta construcción del discurso conforme a mensajes cortos encuentra un espacio fructífero dentro de las redes sociales y sus habituales limitaciones de caracteres relacionadas con el consumo excesivo y la velocidad de lectura que mencionamos

al comienzo del apartado. Como veíamos, ante la desautorización de las distintas instituciones se produce una reconfiguración de la certeza conforme a los propios conocimientos por observación, lo cual requiere de una profunda simplificación del mundo. Precisamente, uno de los atractivos que ofrecen las teorías conspirativas es la confirmación de la propia posición desde un razonamiento superficial, propio de aquel que se promueve en estas plataformas de mensajería rápida. Ejemplo de ello es que durante el desarrollo de una manifestación en el centro de Madrid el pasado mes de octubre, una pancarta rezaba: “si el calor mata, ¿por qué siguen vivos en Arabia Saudí?”.

Esta simplificación de la realidad está sustentada sobre una formación discursiva apo-yada sobre la comunidad, tanto en su dimensión epistémica como en su dimensión personal, donde se cruzan las cuestiones de la integración y el reconocimiento. Esta comunidad conforma, como hemos visto, una burbuja epistémica, sin embargo, se trata tanto de una burbuja algorítmica (Lara 2017), provocada por la recomendación maquina, como de ciertas dinámicas entre los usuarios dentro del grupo. En estos casos, la burbuja epistémica tiene dos causas estrechamente relacionadas: la polifonía y la censura.

La polifonía hace referencia a la multitud de grupos similares donde se encuentran muchos usuarios desde donde se reenvía constantemente información hacia el grupo. Esto genera una red de información a menudo falsa y construye una comunidad epistémica sobre la que se cimantan las creencias a través de estrategias de intertextualidad y delegación de autoría. En este sentido, se podría decir que la burbuja epistémica no se reduce al grupo de estudio sino que se extiende rizomáticamente hacia un centenar de grupos más. Esto se relaciona con el último punto de la decatupla bungiana, según el cual “la pseudociencia no tiene parientes próximos salvo quizá otra pseudociencia con los que pueda interactuar fructíferamente; o sea, la pseudociencia está prácticamente aislada, no existe un sistema de pseudociencias paralelo al de las ciencias” (Bunge 1985, pp. 69-70). A menudo se piensa que las burbujas de información son sólo estos procesos sin ningún tipo de intervención centralizada; sin embargo, en el mismo grupo se generan núcleos de resistencia que quedan censurados por los administradores. A este respecto, F. advertía: “los antiguos organizadores me echaron del grupo por poner en duda su teoría del cambio climático causado por geoingeniería”. El ethos de esta comunidad “lejos de ser el de la libre búsqueda de la verdad, de la profundidad y de la sistematicidad,

es el de la defensa obstinada del dogma, si es necesario con ayuda del engaño o la violencia” (Bunge 1985, p. 70).

En general, el crecimiento del grupo durante los últimos meses hasta rozar los dos mil miembros se construye sobre las mismas estructuras digitales de la aplicación que permiten un control pormenorizado del discurso y una sensación de integración y reconocimiento por parte del resto de miembros. En este sentido, los grupos conspirativos no son una manzana podrida dentro de los ecosistemas digitales, sino en todo caso la versión hiperbólica de los vicios epistémicos propios de las actuales tecnologías de la comunicación y la información.

## 5. Conclusiones y alternativas

Para poder ofrecer una ciencia solvente ante la malversación del conocimiento científico de las teorías de la conspiración debemos asimismo descartar la visión dogmática de la ciencia que domina en gran parte de la sociedad y que la convierte en un conocimiento infalible, definitivo y sobre todo ajeno al contexto social en el que emerge. Así pues, aunque Feyerabend exagera al aseverar que la ciencia no es más que una ideología al mismo nivel que los cuentos de hadas, no yerra en su diagnóstico: “en la sociedad en general el juicio científico es recibido con la misma reverencia con la que no hace mucho tiempo se aceptaba el juicio de los obispos y los cardenales” (Feyerabend 1982, p. 82). No es infrecuente leer titulares periodísticos que anuncian “el último descubrimiento” o “la respuesta definitiva” de la ciencia (referida siempre en singular, como un todo macizo en lo epistémico, lo metodológico y lo sociológico) ante un determinado problema. Esto no hace sino evidenciar la confianza muchas veces irrazonada de la población no conspirativa hacia las instituciones científicas, lo que refuerza la dialéctica entre esta y las personas conspirativas, y alimenta la autoimagen crítica y antidogmática de estos últimos.

Ante esto, la filosofía de la ciencia social y feminista nos ofrece una alternativa: una ciencia plural y diversa, tanto entre los sujetos que la practican como en las metodologías y valores que la guían (Longino 1990). Asimismo, esta perspectiva permite integrar la influencia de factores sociales en la ciencia sin separarla de su dimensión empírica y sus métodos. Tacoronte (2020) dice al respecto de la propuesta de Longino:

Su empirismo es contextual y da un lugar central a la comunidad científica. La reflexión epistemológica tiene como objetivo resaltar la ciencia y su actividad como práctica

social, porque esta autora entiende que la acción científica es intersubjetiva. La actividad científica es contextual, no individualista y pluralista (p. 52).

La experiencia sensible que es la base del conocimiento científico debe incorporarse siempre a un corpus teórico ya establecido. Este corpus estará siempre atravesado por ideologías, valores y prácticas sociales que pueden ser reaccionarias o progresistas y, querámoslo o no, contribuirán al modo en que interpretaremos y explicaremos las evidencias que provienen de los sentidos. De esta forma, la ciencia nunca es neutral ni arroja resultados definitivos, pero tampoco es una fábula que pueda malearse a placer por los sujetos. Los valores sociales deben convivir con los epistémicos y con la evidencia empírica.

Otro referente a este respecto es Philip Kitcher, que considera imprescindible encuadrar la ciencia dentro de la sociedad y sus valores, pero sin caer en lo que denomina “igualdad epistémica”: la idea según la cual las opiniones de todo el mundo valen lo mismo y están igualmente fundadas. En su obra *Science in a Democratic Society* (2011) distingue entre tres tipos de valores a tener en cuenta:

1. Valores amplios: relativos a la libertad, la justicia o la democracia.
2. Valores probativos: que atañen a las líneas de conocimiento que vale la pena investigar y cómo.
3. Valores cognitivos: que expresan un compromiso en adquirir ciertos tipos de conocimiento. precisamos de una relación más fluida, horizontal y bidireccional entre centros de producción de conocimiento científico y sociedad civil.

La ciencia de calidad será aquella que logre conjugar y coordinar estos tres tipos de valores. A su vez, la pseudociencia se caracterizará por su incapacidad para integrar los valores sociales y la crítica que de ellos se derivan sin renunciar a una actitud epistémicamente constructiva y fiel a la experiencia sensible. Frente al escepticismo radical de los conspirativos, pero también frente a cualquier atisbo de confianza ciega en la ciencia, debemos reivindicar un escepticismo moderado. En palabras de Mario Bunge (1985):

mientras el escéptico radical es nihilista, el escéptico moderado es constructivo. Y lo que construye, a diferencia del edificio dogmático, no se desploma al primer temblor porque

ya ha pasado pruebas escépticas [...]. Un buen demócrata es un escéptico moderado: examina y sopesa sin obedecer ni desobedecer de primeras (Bunge 2014, p. 198).

Solo de esta manera estaremos en condiciones de confrontar las posturas conspirativas, no tanto (o no solo) con hechos científicos, sino con una actitud alternativa saludable tanto en lo epistémico como en lo social.

## Referencias

- Brahms, Y. (2020). *Philosophy of Post-Truth*. Institute for National Security Studies, pp. 1-19.
- Bunge, M. (1982). Demarcating Science from Pseudoscience. *Fundamenta Scientiae*, 3 pp. 369-388.
- Bunge, M. (1985). *Pseudociencia e ideología*. Madrid: Alianza Editorial.
- Bunge, M. (2014). *Las pseudociencias ¡vaya timo!*. Barcelona: Laetoli.
- Feyerabend, P. (1982). *La ciencia en una sociedad libre*. Madrid: Siglo XXI Editores.
- Gallo, A. (2019). Teorías de la conspiracionismo: de la paranoia al genocidio. *Estudios humanísticos. Filología*, 41, pp. 217-243. doi: 10.18002/ehf
- García, J. (2018). La posverdad en la difusión de la información científica. *La posverdad y las noticias falsas: el uso ético de la información*, pp. 177-201.
- Hansson, S. (2021). Science and Pseudo-Science. *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/fall2021/entries/pseudo-science/>. Consulta: 09/10/2023.
- Kitcher, P. (2011). *Science in a Democratic Society*. Londres: Prometheus Books.
- Lakatos, I. (2018). La historia de la ciencia y sus reconstrucciones racionales. En Hacking, I. (ed.) *Revoluciones científicas* (pp. 204-242). Ciudad de México: Fondo de Cultura Económica.

- Lara, P. (2017). 'Homo Facebook': ventanas conceptuales, posverdad y burbuja algorítmica. *COMeIN*, 71. doi: <https://doi.org/10.7238/c.n71.1767>
- Laudan, L. (1983). The demise of the demarcation problem. En Cohen, R. *Physics philosophy and psychoanalysis*. Boston: Library of Congress Cataloging in Publication Data.
- Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.
- Meyer, M., Oeberst, A., & Imhoff, R. (2022). How Do Conspiratorial Explanations Differ from Non-Conspiratorial Explanations? A Content Analysis of Real-World Online Articles. *European Journal of Social Psychology*. doi: 10.1002/ejsp.2903
- Popper, K. (1962). *Conjectures and refutations. The growth of scientific knowledge*. Nueva York: Basic Books.
- Popper, K. (1994). *La sociedad abierta y sus enemigos*. Barcelona: Paidós.
- Ramonet, I. (2023). *La era del conspiracionismo. Trump, el culto a la mentira y el asalto al Capitolio*. Madrid: Siglo XXI Editores.
- Riechmann, J. (2020). La crisis del coronavirus y nuestros tres niveles de negacionismo. *The Conversation*. <https://theconversation.com/la-crisis-del-coronavirus-y-nuestros-tres-niveles-de-negacionismo-134749>. Consulta: 09/10/2023.
- Tacoronte, M. (2020). Helen Longino. Una epistemología empirista y pluralista. *Cuestiones de género: de la igualdad y la diferencia*, 15 , pp. 51-70.
- Urteaga, E. (2013). Percepción estudiantil de la ciencia y tecnología en el País Vasco. *Sociología y Tecnología. Revista Digital de Sociología del Sistema Tecnocientífico*, 4(1) , pp. 58-74.
- Zafra, R. (2017). Redes y posverdad. En Fanés, J. y Maldonado, M. (eds.). *En la era de la posverdad*. Madrid: Calambur.



# Procesamiento predictivo y bayesianismo: una brevísima aproximación filosófica-probabilística al modelamiento predictivo de la mente

Lucas Hinojosa-López<sup>a</sup>

## Resumen

El procesamiento predictivo es una de las tantas aproximaciones estadísticas a la comprensión del funcionamiento de la mente de los últimos quince años. El modelamiento estadístico en la comprensión del cerebro y la mente son comprensiones normativas, en tanto que proponen que las funciones de la mente operan en conjunto con las normas de las probabilidades. El procesamiento predictivo sería entonces una forma de modelar probabilísticamente las funciones mentales, en donde el cálculo de las probabilidades y su análisis filosófico cumplen un rol interesante de analizar. El objetivo de este trabajo es doble: en primer lugar (1) se busca presentar el procesamiento predictivo como una aproximación filosófico-estadística al modelamiento de la mente; y en segundo lugar (2) se analizará el rol que cumple el bayesianismo en dicha evaluación probabilística-estadística, en el que las reglas bayesianas explicarían, al menos parcialmente, las distintas operaciones predictivas de la mente

**Palabras Clave:** Procesamiento predictivo, Regla de Bayes, Probabilidad, Creencia, Mente.

## 1. Introducción

El procesamiento predictivo (o *predictive processing* en inglés) es un marco teórico en ciencias cognitivas y neurociencias que sugiere que

---

<sup>a</sup>Universidad de Valparaíso, Valparaíso, Chile.  
Contacto: lucas.hinojosa@uv.cl

el cerebro humano no solo reacciona a la información sensorial que recibe del entorno, sino que también genera activamente predicciones sobre esa información y las utiliza para interpretar y dar sentido a su entorno. Lo que propone el marco del procesamiento predictivo es que la percepción está continuamente construyéndose a través de un ciclo continuo de predicciones y correcciones basadas en la comparación entre las expectativas del cerebro y las entradas sensoriales reales (Friston 2005).

Es posible comprender el procesamiento predictivo (PP de aquí en adelante) como una aproximación estadística a la comprensión del cerebro y la mente (e.g Hohwy 2013; Clark 2016) en donde las acciones o formulaciones de hipótesis realizadas por agentes en contextos de desenvolvimiento habituales serían producto de un cálculo interno probabilístico del cerebro y la mente. Se espera a partir de este marco que todas las funciones y disfunciones mentales se puedan entender como respondiendo al principio de minimización de error de predicción. Según este principio, el cerebro ajusta constantemente sus modelos internos para reducir la discrepancia entre las predicciones generadas y las entradas sensoriales reales (Friston 2005). Este proceso cíclico mejora y reajusta la precisión de las predicciones y optimiza la adaptación perceptual. Hablaremos de este principio más adelante.

El bayesianismo tiene un lugar importante en este marco teórico ya que el ajuste constante del PP radica en la naturaleza probabilística de las predicciones y la actualización de las creencias. El cerebro, al predecir y comparar constantemente las expectativas con las entradas sensoriales reales, sigue un marco bayesiano. La teoría de Friston (2010) conecta el procesamiento predictivo con la inferencia bayesiana, donde el cerebro formula y actualiza constantemente creencias probabilísticas para minimizar la incertidumbre. Este enfoque bayesiano es importante en términos predictivos y probabilísticos, pues ayuda a explicar cómo un modelo mental interno actual podría ajustarse para mejorar la precisión predictiva en entornos dinámicos. El modelamiento bayesiano puede encontrar un territorio común al procesamiento predictivo debido a este poder unificador que brinda el principio de minimización de error de predicción. En general, se espera que el PP pueda dar cuenta del rol agencial de un agente a partir del análisis probabilístico y estadístico del reajuste de hipótesis que nos permiten actuar en el mundo.

A continuación, presentaremos más en detalle el enfoque del PP, su utilidad para la comprensión del funcionamiento de la mente y qué relación tiene con el cálculo de probabilidades en el proceso predictivo.

Defenderemos que el bayesianismo tiene un rol importante para el PP en tanto que explica la actualización de creencias e hipótesis que nos permiten actuar.

## 2. Procesamiento Predictivo (PP)

Como ya se ha adelantado previamente, el procesamiento predictivo representa una perspectiva revolucionaria en la neurociencia cognitiva y que desafía la visión convencional de cómo el cerebro interactúa con su entorno (Clark 2016). Lo anterior se debe a la idea de que el cerebro no es solo un receptor pasivo de información sensorial, sino que constantemente genera predicciones activas sobre el mundo que le rodea (Friston 2005; Hohwy 2013). En este sentido, la teoría del procesamiento predictivo propone que la percepción no es simplemente el resultado de una reacción a estímulos externos, es decir, no opera mediante la mecánica de *inputs* y *outputs*, sino más bien mediante el resultado de una construcción activa de modelos internos que anticipan y explican eventos futuros.

El enfoque del PP encuentra sus raíces en el trabajo seminal de Karl Friston “A theory of critical responses” (2005) donde formuló sus bases teóricas. Friston propuso el principio de “energía libre” argumentando que el cerebro opera para minimizar la sorpresa o la discrepancia entre sus predicciones internas y las entradas sensoriales reales, cuestión que podríamos llamar incertidumbre. Este principio impulsa un ciclo continuo de ajuste donde el cerebro actualiza constantemente sus modelos internos para optimizar la precisión predictiva y reducir la incertidumbre.

El PP se basa en la premisa de que el cerebro construye activamente su realidad perceptual a través de la anticipación y corrección constante. Durante este proceso, el cerebro formula predicciones sobre los estímulos sensoriales futuros basándose en sus experiencias pasadas y en los patrones aprendidos. Estas predicciones son comparadas con las entradas sensoriales reales y cualquier discrepancia resultante se utiliza para ajustar y refinar modelos internos. De esta forma, el cerebro no solo reacciona al entorno (no solo responde a *inputs*), sino que también se anticipa a él, transformando la percepción en un acto dinámico y proactivo (Parr et al. 2022).

El principio de minimización del error de predicción es de suma relevancia en este contexto. Este principio establece que el cerebro ajusta sus modelos internos para reducir la discrepancia entre las predicciones

generadas y las entradas sensoriales reales. La minimización del error de predicción se convierte en una herramienta clave para mejorar la adaptación perceptual y optimizar la capacidad del cerebro para predecir eventos futuros. El trabajo de Friston (2010) extiende esta idea al principio de energía libre, que formaliza cómo el cerebro minimiza la incertidumbre y optimiza sus creencias probabilísticas para explicar y predecir el mundo.

Imaginemos que caminamos por un bosque y escuchamos un sonido crujiente detrás de unos arbustos. Inevitablemente, generaremos una predicción sobre la posible fuente de dicho sonido: digamos en este caso que la fuente del sonido es un animal. Esta predicción se construye a partir de experiencias previas y en modelos internos construidos a lo largo del tiempo en los que mi experiencia me indica que resulta lógico (i.e., esperable en función al contexto) para el lugar en el que estoy, que un conejo o quizás un perro haya pisado hierba seca o ha movido un arbusto.

Decidimos acercarnos al arbusto para indagar e intentar dilucidar nuestra interrogante. Resulta que el sonido es causado por el viento moviendo ramas secas. En este punto hay una discrepancia entre la predicción inicial del cerebro (i.e., la presencia de un animal) y la realidad percibida (el viento moviendo las ramas). El cerebro y nuestra mente utilizan estas discrepancias o errores predictivos para reajustar los modelos internos.

Esto sugiere, según el principio de minimización de error de predicción (Friston 2010), que cuando escuchemos un sonido similar en un medio boscoso estaremos mejor preparados para predecir que el sonido podría ser causado por el viento en lugar de un animal. Este proceso iterativo de hacer predicciones, enfrentarse a la realidad y ajustar modelos internos es necesario para minimizar el error de predicción y mejorar la adaptación perceptual.

El nexo entre el procesamiento predictivo (e.g., Hohwy 2013) y la inferencia bayesiana (e.g., Hartigan 1983) revela una conexión profunda en la forma en que el cerebro aborda la incertidumbre y actualiza sus creencias. La teoría del PP desarrollada por K. Friston se asocia estrechamente con los principios fundamentales de la inferencia bayesiana. La relación entre ambos marcos teóricos resalta cómo el cerebro no solo predice activamente, sino que también utiliza la información actual y previa para ajustar sus creencias y reducir la incertidumbre. Abordaremos esto en el apartado siguiente.

### 3. El rol de la inferencia bayesiana en el procesamiento predictivo

La inferencia bayesiana es un enfoque estadístico que se basa en la actualización de creencias a medida que se acumula nueva evidencia. Este proceso refleja la idea de que el conocimiento es inherentemente incierto y que la certeza se alcanza mediante una actualización continua basada en información disponible (Earman 1992). En el contexto del PP, la inferencia bayesiana se convierte en la piedra angular que fundamenta cómo el cerebro optimiza sus modelos internos y ajusta sus creencias.

Como se mencionó anteriormente, autores como Friston (2005; 2010) han propuesto que el cerebro opera para minimizar la incertidumbre, la discrepancia entre las predicciones internas y las entradas sensoriales reales, a través de un constante reajuste probabilístico en las creencias del agente para mantenerse en sintonía con el medio y reducir la incertidumbre.

En el marco de la inferencia bayesiana, las creencias previas se combinan con la nueva evidencia a través de la regla de Bayes (1764) para actualizar las creencias posteriores. Dicha regla es un principio fundamental en teoría de la probabilidad, pues permite actualizar nuestras creencias sobre un evento en función de la evidencia disponible (Easwaran 2011). La regla de Bayes en su versión más simple se obtiene directamente de la probabilidad condicional y el axioma de multiplicación. Se expresa de la siguiente manera:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

En este caso,  $P(B)$  corresponde solo a una hipótesis y su complemento que dan la probabilidad total. Sin perjuicio de lo anterior, debe destacarse que lo interesante de la regla es, al menos en el caso de su versión simplificada, calcular la probabilidad total sobre la probabilidad de todas las hipótesis posibles y de las respectivas verosimilitudes.

En esta regla (Parzen 1979) tenemos que:

1.  $P(A|B)$  es la probabilidad de que el evento A ocurra dado que B ha ocurrido; es decir, esta es la probabilidad posterior o inversa de A.
2.  $P(B|A)$  es la probabilidad de que B ocurra dado que A ha ocurrido.

3.  $P(A)$  y  $P(B)$  son las probabilidades iniciales de que A y B ocurran, respectivamente.

La probabilidad posterior de determinado evento es producto de (i) la probabilidad anterior y (ii) del *Prior* y de la verosimilitud de la evidencia dada la hipótesis, es decir, A. La probabilidad anterior (aquello a lo que se le suele llamar *Prior*) es la hipótesis que resulta más probable independientemente de cualquier dato que ingrese al sistema interno del agente. Dado el contexto o situación en la que podamos encontrarnos, hay ciertas hipótesis más probables que otras. Por ejemplo, si estoy haciendo clases y siento algo tocando la puerta, resulta más probable que sea un estudiante llegando tarde a la clase en lugar de un pájaro carpintero golpeando la madera de dicha puerta en la región de Valparaíso. Si bien las dos probabilidades podrían explicar el fenómeno, una hipótesis es más probable que otra dado el contexto y porque hay probabilidades anteriores que son mayores: en este caso, la probabilidad de que un estudiante llegue tarde a la clase. Análogamente, el cerebro en el procesamiento predictivo genera predicciones basadas en creencias internas previas y las ajusta a medida que llega nueva información. La adaptación continua refleja la naturaleza bayesiana de cómo el cerebro actualiza sus modelos internos y mejora la precisión predictiva.

La aplicación de la inferencia bayesiana en el procesamiento predictivo se refleja justamente en la formulación de hipótesis y la evaluación de la probabilidad de eventos futuros. En este sentido, el cerebro no solo responde a entradas sensoriales, sino que también utiliza la información previa para calcular la probabilidad de distintos escenarios y generar predicciones más precisas. La hipótesis que tiene la probabilidad posterior mayor es la que finalmente gana la competencia por aparecer en la consciencia (Hohwy 2013): este es el precepto que se forma. Lo anterior supone que la hipótesis que se nos aparezca en primer lugar será la hipótesis con la probabilidad posterior mayor. Esto le permitirá al sistema minimizar el error de predicción en el largo plazo.

El cerebro opera típicamente en condiciones de incertidumbre (Clark 2016). Si hay ruido en el mundo, el sistema debe inferir qué tan fiable es la señal que percibe. La señal sensorial puede ser fiable o no dependiendo del contexto. Se debe tener expectativas respecto de la precisión o la fiabilidad de la señal sensorial, pues si no las hay, no hay forma razonable de evaluar cuánto ajuste necesitan nuestras creencias en el contexto, y la operación bayesiana ayuda justamente a reajustar la probabilidad que le asignamos a las distintas variables. Esto implica que, si bien busca minimizarse el error de predicción, este es necesario pa-

ra saber cuánto ajuste requieren nuestras hipótesis sobre determinado evento o fenómeno.

Lo anterior nos lleva a la cuestión de la evidencia acumulativa (Easwaran 2011), pues la inferencia bayesiana también se aprecia en la consideración de la evidencia acumulativa. En la inferencia bayesiana, la evidencia se suma de manera incremental para actualizar las creencias. De manera similar, en el PP cada iteración de ajuste se basa en la acumulación de experiencias pasadas y en la incorporación de nuevas observaciones. Las inferencias perceptuales que podemos realizar deben ser desarrolladas en función de expectativas relativas a la precisión del error de predicción. Esto nos permitiría entonces en algunos casos, a nivel interno y subjetivo, modelar la precisión del mundo. Las precisiones determinan qué tanta influencia tienen mis hipótesis sobre el mundo. Si se determina que cierta señal no es fiable, entonces las hipótesis guiarán mi comportamiento o razonamiento. Si la señal y los datos que ingresan al sistema son por otro lado fiables, entonces se le da mayor ganancia a ese error de predicción en el momento de revisión de hipótesis. Estos errores ajustan nuestros perceptos sobre el mundo.

La conexión entre el procesamiento predictivo y la inferencia bayesiana revela una sinergia en la forma en que el cerebro aborda la incertidumbre y actualiza sus creencias. La teoría de Friston (principalmente apoyándonos en sus escritos de 2005 y 2010) ofrece una perspectiva unificadora que destaca la naturaleza probabilística del cerebro y su capacidad para adaptarse en un entorno dinámico. Esta perspectiva probabilista puede explicarse en términos bayesianos (Earman 1992; Easwaran 2011; Hartigan 1983) de la forma en que aquí ha sido expuesta.

#### **4. Lucha de probabilidades: formas de reducir el error de predicción**

El procesamiento predictivo se encarga de asignar probabilidades a eventos futuros en un intento de predecir fenómenos o situaciones, lo que permite conjeturar hipótesis y de esta manera evitar la incertidumbre de determinado agente. Según Ramsey (1926, como se menciona en Ramsey 1931) la probabilidad debe ser definida como grado de creencia. Las hipótesis que se juegan en PP se confirman bayesianamente, lo que indica que en línea con Ramsey dicha confirmación o desconfirmación de hipótesis se realiza a través de una comparación de los grados de creencia que se tienen en determinado evento (Howson y Urbach 1993).

Mientras más probable sea la evidencia relativa a la hipótesis conjeturada en el procesamiento predictivo, más confirmada (i.e., más grado de confirmación) será esta última (Howson y Urbach 1993, p. 93).

En línea con la regla bayesiana, la probabilidad posterior o resultante de una hipótesis dependerá de la probabilidad anteriormente asignada. Esto sugiere que una hipótesis se confirma cuando su probabilidad posterior excede su probabilidad anterior y mientras mayor sea la diferencia, mayor será la confirmación de ella (Howson y Urbach 2006, p. 100).

Esto es sumamente relevante en el contexto del procesamiento predictivo en donde el cerebro y la mente cometen de forma inevitable errores predictivos y el cerebro debe reajustar el sistema de modelado interno para incorporar nuevas variables en el cálculo probabilístico de la predicción. Aquí es donde inician las luchas de probabilidades internas para reajustar la predicción y evitar los errores predictivos. Ejemplifiquemos esto de la siguiente manera, relacionando el procesamiento predictivo con la regla de Bayes y asignando valores:

Imaginemos un entorno en el que generalmente encontramos animales. Nuestro cerebro ha desarrollado modelos internos para predecir la probabilidad de encontrarnos con ciertos tipos de animales. Caminando por un bosque se escucha un sonido a la distancia. A partir de experiencias previas realizamos una predicción inicial: existe un 70 % de probabilidad de que el sonido haya sido producido por un pájaro, debido a nuestros recuerdos sobre cómo suena un pájaro, un 20 % de que sea un mamífero y un 10 % de que sea otro tipo de animal. Estas últimas dos estimaciones son considerablemente bajas dada nuestra fuerte creencia de que el sonido proviene de un pájaro y que a su vez el sonido no es proveniente de un mamífero u otro.

Al aplicar la regla de Bayes a este ejemplo se deben actualizar estas probabilidades en función de nueva información. Supongamos entonces que observamos un movimiento en los arbustos y logramos identificar cierto patrón de plumaje: la evidencia sugiere fuertemente que el sonido ha sido producto de un pájaro. La regla de Bayes nos permite ajustar las probabilidades iniciales para reflejar esta nueva información. Se presenta a continuación:

Si aplicamos la regla de Bayes

$$P(\text{Sonido de Pájaro}) = \frac{0,95 \cdot 0,70}{0,95 \cdot 0,70 + 0,03 \cdot 0,20 + (0,02 \cdot 0,10)}$$



<b>PROBABILIDAD INICIAL</b>	<b>AJUSTE A PARTIR DE EVIDENCIA</b>
Pájaro (Estimación a partir del sonido emitido por el objeto): 70 %	Pájaro (Estimación dada la verosimilitud entre Patrón de plumaje/sonido/pájaro): 95 %
Mamífero: 20 %	Mamífero: 3 %
Otro: 10 %	Otro: 2 %

Cuadro 1: Ajuste de probabilidades basado en nueva evidencia.

Después de realizar estos cálculos podríamos encontrar una nueva probabilidad ajustada de que el sonido sea en efecto un pájaro. Este proceso refleja cómo la regla de Bayes permite al cerebro actualizar sus creencias iniciales en función de nueva evidencia. Esta nueva probabilidad podría ser, por ejemplo, la siguiente: Pájaro 80 %, Mamífero 15 % y otro animal 5 %. Resultaría finalmente algo así:

<b>PROBABILIDAD INICIAL</b>	<b>AJUSTE A PARTIR DE EVIDENCIA</b>	<b>NUEVA PROBABILIDAD</b>
Pájaro: 70 %	Pájaro: 95 %	Pájaro: 80 %
Mamífero: 20 %	Mamífero: 3 %	Mamífero: 15 %
Otro: 10 %	Otro: 2 %	Otro: 5 %

Cuadro 2: Evolución de las probabilidades tras el ajuste bayesiano.

Este proceso se repite cada vez que el sistema adquiere nueva información y esta pueda ser utilizada para modelar las predicciones y, por lo tanto, conjeturar hipótesis sobre lo que nos rodea. Cada ajuste incremental refleja cómo el cerebro utiliza la regla de Bayes para reducir el error de predicción actualizando sus modelos internos en base a la evidencia acumulativa. La probabilidad final se ajusta continuamente a medida que el cerebro enfrenta nuevas situaciones, mejorando así su capacidad para predecir y adaptarse al entorno. La aplicación de la regla de Bayes destaca cómo el cerebro produce, actualiza y refina las inferencias inductivas utilizadas para conjeturar hipótesis: el cerebro no solo predice, sino que también aprende y ajusta sus creencias en función de la información disponible.

Este enfoque probabilístico y adaptativo es de suma relevancia para reducir el error de predicción y optimizar la capacidad del cerebro para anticipar y comprender su entorno en constante cambio.

## 5. Consideraciones finales

A lo largo de este escrito hemos hablado de conjeturas e hipótesis, cuestiones que refieren principalmente a los razonamientos inductivos. Las inducciones que se realicen dependen de experiencia previa y las conjunciones de experiencias semejantes del pasado (Pereira 2009, p. 113) y a su vez de la suposición de que el futuro será como el pasado: esto es lo que permite la predicción.

Las predicciones que se desprenden del PP califican como inducciones y razonamientos ampliativos al ofrecer nuevas probabilidades de que determinado evento, suceso o fenómeno resulte verdadero en función de la actualización de evidencias.

No hemos abordado la naturaleza del procesamiento predictivo en términos de experiencia fenomenológica -o experiencia consciente o inconsciente- pues no ha sido el objetivo de este escrito. No obstante, sabemos que amerita un análisis más profundo. Adelantamos que algunos autores llaman ‘inferencia perceptual’ a la predicción inconsciente e ‘inferencia activa’ a la inferencia consciente, es decir, la predicción que se torna consciente al aparecer un precepto sobre el cual analizar probabilísticamente la ocurrencia de un evento.

Además, y en relación con lo anterior, debemos aclarar que las probabilidades o porcentajes en el ejemplo ofrecido están sobresimplificados dada toda la información que supone el ejemplo. El *prior* por ejemplo es un valor que debería aclararse, pero dado el objetivo de este escrito -que no tiene por finalidad realizar demostraciones matemáticas o lógicas- esa información no se entrega. Para tales fines, lo óptimo y pertinente sería utilizar la fórmula completa de Bayes y aquí solo hemos utilizado la versión simplificada.

Por último, debe explicitarse que el enfoque bayesiano sugiere que las probabilidades pueden condicionarse por otras probabilidades en función de eventos previos: este fue el caso del sonido del pájaro en donde el avistamiento de plumaje aumentaba el grado de creencia de que el animal en cuestión fuese un pájaro. En un paradigma bayesiano, y más aún bajo el marco del procesamiento predictivo, la confirmación de las hipótesis resulta ser una cuestión gradual: esto nos llevará al condicionante de que si la probabilidad de evidencia sube entonces

también lo hará la probabilidad de confirmación de la hipótesis. Lo anterior no ocurriría solo en los análisis lógicos-estadísticos sino también en las predicciones internas del cerebro, la mente y los modelos internos del aparato mental. El bayesianismo recoge estos aspectos de no monotonicidad o contextuales y, sin embargo, los encaja en un esquema inferencial probabilístico que refleja cómo modificamos nuestras creencias y aprendemos de la experiencia.

## Referencias

- Bayes, T. (1764). An Essay Toward Solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, 53 pp. 370-418.
- Clark, A. (2016). *Surfing uncertainty*. <https://doi.org/10.1093/acprof:oso/9780190217013.001.0001>
- Earman, J. (1992). *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory*. MIT Press.
- Easwaran, K. (2011). Bayesianism I: Introduction and arguments in favor. *Philosophy Compass*, 6(5) pp. 312-320. <https://doi.org/10.1111/j.1747-9991.2011.00399.x>
- Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B*, 360(1456) , pp. 815-836. <https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2) , pp. 127-138. <https://doi.org/10.1038/nrn2787>
- Hartigan, J. (1983). *Bayes Theory*. Springer series in statistics. <https://doi.org/10.1007/978-1-4613-8242-3>
- Hohwy, J. (2013). *The predictive Mind*. <https://doi.org/10.1093/acprof:oso/9780199682737.001.0001>
- Howson, C., & Urbach, P. (1993). Scientific reasoning: the Bayesian approach. *Journal of the Royal Statistical Society*, 156(1), p. 130. <https://doi.org/10.2307/2982875>
- Howson, C., & Urbach, P. (2006). *Scientific reasoning: The Bayesian approach*.

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The Free Energy Principle in Mind, Brain, and Behavior*. MIT Press.

Parzen, E. (1979). *Teoría moderna de probabilidades y sus aplicaciones*.

Pereira, F. (2009). *David Hume: naturaleza, conocimiento y metafísica*.

Ramsey, F.P. (1926). "Truth and Probability" in *The Foundations of Mathematics and other Logical Essays*, Ch. VII, p. 156-198, edited by R.B. Braithwaite, London: Kegan Paul, Trench, Trubner & Co. New York: Harcourt, Brace and Company.

# Teorías humeanas y No humeanas: el desacuerdo con respecto a los inobservables

Kuminak Lefio Zamorano<sup>a</sup>

## Resumen

El adscribir a una posición humeana o no humeana es una decisión importante al proponer una teoría metafísica. Esta decisión resulta especialmente relevante cuando nos preguntamos acerca de qué justifica la postulación de inobservables en esta área de la filosofía. Los inobservables son entidades acerca de las cuales no tenemos una experiencia directa, sino que postulamos su existencia por otras razones, generalmente por un interés en explicar nuestra experiencia del mundo y sus regularidades. Este ensayo se sitúa ante este desacuerdo específico, intentando dar una caracterización general de ambas posturas y analizar lo que cada postura estaría considerando valioso, y valioso para que, al defender su perspectiva con respecto a la postulación de inobservables. Esto implica un análisis más profundo, sobre cómo podemos discutir racionalmente acerca de aquellos aspectos que valoramos negativa o positivamente en la conformación de nuestras propuestas teóricas y bajo qué meta-teoría sobre la justificación de nuestras teorías filosóficas abordamos estas preguntas. Me centraré inicialmente en la revisión del problema de los inobservables entre teorías humeanas y no humeanas, para luego mostrar los problemas más generales que surgen al preguntarnos por la justificación de nuestras posturas filosóficas.

**Palabras clave:** Valores, normatividad, racionalidad, punto de vista.

## Abstract

Adherence to a Humean or non-Humean position represents a significant choice when proposing a metaphysical theory. This decision is

---

<sup>a</sup>Universidad de Chile, Santiago, Chile.  
Contacto: gklefio@gmail.com

especialmente relevante cuando nos preguntamos sobre lo que justifica la postulación de unobservables en esta área de filosofía. Unobservables son entidades sobre las que no tenemos experiencia directa, sin embargo, postulamos su existencia por otras razones, generalmente porque de un interés en explicar nuestra experiencia del mundo y sus regularidades. Este ensayo se sitúa frente a esta específica discrepancia, tratando de caracterizar ambas posturas en general y analizar lo que cada postura debería considerar valioso, y valioso para qué, cuando defendemos nuestra perspectiva sobre la postulación de unobservables. Esto implica un análisis más profundo, sobre cómo podemos racionalmente discutir sobre esos aspectos que valoramos negativamente o positivamente en la formación de nuestras propuestas teóricas y bajo qué meta-teoría sobre la justificación de nuestras teorías filosóficas abordamos estas cuestiones. Inicialmente, me centraré en el problema sobre los unobservables entre teorías humeanas y no-humeanas y luego mostraré los problemas más generales que surgen cuando nos preguntamos sobre la justificación de nuestras posiciones filosóficas.

**Keywords:** Valores, normatividad, racionalidad, punto de vista.

## 1. Introduciendo las intenciones del trabajo

Al adentrarnos en la discusión metafísica acerca de las leyes de la naturaleza y, de forma más general, acerca de las regularidades que percibimos y experimentamos en nuestro paso por el mundo, nos encontramos con dos enfoques que, sin ser los únicos posibles, han marcado históricamente un choque de perspectivas que se mantiene relevante hasta hoy en día. Sin ser necesario que todo el debate pueda reducirse a esta discusión, sí vale la pena notar cómo tomar una postura respecto a los desacuerdos entre una teoría humeana (TH) y una teoría no humeana (TNH) es una decisión importante al entablar una discusión metafísica sobre estas temáticas. Considero que es un ejemplo de lo que veremos más adelante como TH las propuestas de Lewis (1986; 1994) y un ejemplo de una TNH lo propuesto por Armstrong (1983); espero que esto sirva para ilustrar el tipo de posturas que estoy considerando, aunque el enfoque de mi trabajo se trata de reconstruir directamente lo que entiendo que defiende cada postura filosófica en general y no según ejemplos específicos. Como guía en esta clasificación de posturas metafísicas estoy aplicando lo propuesto por Soto. C. (2024).

La importancia de distinguir entre adoptar una postura afín a una TH o TNH se justifica en cuán fundamentales son las cuestiones que separan los diversos enfoques inspirados por cada una de ellas. Lo que delimitará aquello que será una postura afín a una TH o a una TNH

serán el tipo de respuestas que cada postura dé ante una serie de problemas filosóficos. Esta misma idea justifica la estructura de este trabajo, es decir, centrarnos en un problema metafísico específico y caracterizar las TH y TNH según cómo responden al problema y las posibles justificaciones que apoyan estas respuestas divergentes entre sí. Existen una multiplicidad de puntos de desacuerdo que podrían satisfacer lo antes dicho; yo me terminé decantando por analizar la decisión acerca de cuándo realizar postulados metafísicos y específicamente sobre la postulación de inobservables. La concepción de inobservables que utilizaremos será una conceptualización bien básica: las entiendo como entidades que postulamos al intentar entender la conformación de nuestra realidad y que no podemos experimentar; por ende, tampoco verificamos su adecuación a esta realidad por el medio más directo de la experiencia, en cambio las postulamos por algún tipo de utilidad (generalmente utilidad explicativa) y justificamos su adecuación por otros medios. Esta conceptualización deja mucho por precisar, por ejemplo, con respecto a qué se entendería como una experiencia directa y cuán mediada o qué tipos de mediaciones entre entidades y nuestra experiencia nos llevarían a considerarlas “inobservables” y cuáles contarían simplemente como una extensión de nuestra capacidad de observación. Sin embargo, el debate más general entre las TH y TNH se ha desarrollado sin tanta atención a este tipo de precisiones, sino que apoyándose en una concepción más general de lo que es un inobservable. Así que, aun si es que creo que una precisión mayor de esta conceptualización podría resultar útil para aclarar ciertos desacuerdos, ese no es un objetivo que este ensayo pretenda abarcar. Dicho lo anterior, sí considero necesario en la siguiente sección de este trabajo otorgar una explicitación lo más clara posible acerca de cuál sería este problema acerca de los inobservables y una posible forma de explicar cómo llegamos a esta problemática.

## 2. El problema acerca de los inobservables

Partimos de la idea de que como seres humanos tenemos experiencia de una variedad de cosas, utilizando esta palabra en un sentido muy amplio, lo suficiente justamente para poder abarcar la variedad de tipos de experiencias diferentes con los que nos topamos. Me refiero a la experiencia que tenemos por medio de los diversos sentidos que poseemos, de objetos, de cambios (y por ende los procesos que delimitan nuestra comprensión de estos), de patrones, de otros seres más pareci-

dos o diferentes a mí, de ilusiones, recuerdos, proyecciones, emociones e incluso de nuestra propia conciencia, entre muchas otras experiencias que convergen en la idea general de: tenemos experiencia del mundo por diversos medios. Pero dentro de esta experiencia del mundo, en el sentido general antes planteado, también tenemos la experiencia de nuestro propio interés inquisitivo y también del de otros seres vivos (al menos según cómo los vemos actuar o lo que ellos nos comunican al respecto). Con esto me refiero a la idea básica de que no solo experimentamos el mundo, sino que queremos comprenderlo, tenemos preguntas respecto a cómo es, por qué es de cierta forma o cuál es la naturaleza de lo que experimentamos (entre innumerables más), y estas preguntas no nos exigen tan solo ser respondidas de cualquier forma, nos exigen que se respondan de buena forma, bajo lo que determinemos que son buenas razones y que construyan una explicación convincente de nuestra experiencia del mundo. Por el tipo de seres que somos, no solo experimentamos el mundo, sino que intentamos explicarlo y hacer esto de una buena forma.

Ahora, qué significa cada uno de los puntos de esta última frase es materia de discusión filosófica y en ocasiones metafísica. En el caso del problema que nos atañe en este ensayo, acerca de los inobservables, creo que surge del punto sobre ¿qué significa y delimita una buena explicación de nuestra experiencia del mundo? O lo mismo, pero enfatizado de otra forma: ¿qué define una explicación válida/satisfactoria? Desde esta pregunta podemos ya directamente pasar al problema, separado en dos preguntas acerca de los inobservables: ¿es la explicación de nuestra experiencia a través de la postulación de los inobservables una explicación válida? (Pregunta por la validez). Y desde el otro lado: ¿es lo suficientemente satisfactoria una explicación de nuestra experiencia que no postule inobservables en absoluto? (Pregunta por la suficiencia).

Con lo anterior hemos establecido el problema y podemos pasar a caracterizar las primeras dos respuestas típicas delimitadas según los nombres TH y TNH.

### **3. Explicitando el desacuerdo superficialmente**

Las diferencias entre las TH y las TNH con respecto a “postular inobservables” se pueden expresar de la siguiente forma:

1. Las TH, en general, tenderían a delimitar el alcance de nuestra capacidad de conocer el mundo hasta donde nos lo permite nuestra experiencia, por lo que deciden no postular inobservables,



limitando sus conceptualizaciones a aquello que experimentamos. Desde este punto, pretenden que todos sus postulados metafísicos puedan ser verificados directamente por nuestra experiencia empírica de estos y no pronunciarse más allá de esto.

2. Las TNH proponen la postulación de inobservables en busca de explicar cómo debería ser el mundo para que lo experimentemos de la forma específica en que se nos presenta. Es decir, se postulan conceptualizaciones que no son estrictamente verificables por la experiencia, pero que coinciden con esta en tanto nos permiten aventurarnos a explicar cómo debería ser el mundo al que no accedemos según cómo es lo que sí experimentamos. Desde este punto, buscan justificar y argumentar estos postulados por vías que no serían únicamente la verificación empírica.

Al preguntarnos por qué se produce el desacuerdo, podríamos inicialmente pensar lo siguiente:

- Las TH creen que es lo posible y/o lo más correcto crear nuestras teorías metafísicas basados en la información que obtenemos a través de la experiencia y conceptualizando solo desde esta.
- Las TNH creen que es posible y lo más correcto crear nuestras teorías metafísicas basados en la información que obtenemos a través de la experiencia y con las conceptualizaciones y postulados metafísicos que se necesiten para explicar esta.

Habiendo establecido esto, me interesa analizar esta primera forma de plantear el desacuerdo e indagar en cuáles serían las razones para que este se produzca y, aún más importante, si existen distintos tipos de razones, algunas más bien teóricas (haciendo referencia a los hechos, la evidencia y argumentaciones que se justifican por esta) o evaluativas en un sentido general (haciendo referencia a valores y/o intereses). Siendo mi foco el mostrar cómo estas últimas juegan un rol muy fuerte en la discusión.

Ambas posturas directa o indirectamente responden a la pregunta por la validez de postular inobservables y por la suficiencia de una teoría que carece de ellos. Por un lado, las TH responden negativamente a la validez de los postulados que van más allá de nuestra experiencia directa del mundo, es esto lo que realmente se propone al reducir lo “posible”, no es una afirmación de lo que de hecho podemos hacer porque esto es fácilmente contrastable con la realidad en que vivimos, los

mismos defensores de las TNH son un contraejemplo por sus prácticas metafísicas. Si fuese un hecho la imposibilidad de hacer lo que las TNH proponen, entonces estarían obligados a hacer lo que las TH proponen, no habría desacuerdo por el que discutir si se tratase de una imposibilidad. Este “posible” no tiene que ver con una posibilidad fáctica y punto, sino que con la posibilidad con mantención de la validez; es decir, es imposible realizar una teoría con inobservables que sea válida según las TH.

Con respecto a la pregunta de la suficiencia, las TH tienen aparentemente la opción de contestar tanto afirmativa como negativamente. Sin embargo, considero que el único camino que conserva la coherencia de la postura es que se responda de forma afirmativa. Esto se debe a que la otra posibilidad implica negar nuestra capacidad de explicar el mundo con validez y/o suficiencia de cualquier tipo, es decir, negar absolutamente que la metafísica pueda ser una empresa cognoscitiva valiosa; un escepticismo que termina destruyendo a la misma disciplina por la cual llega a sus propias conclusiones (examinando esto con más tiempo se vuelve visible que una postura de este estilo negaría su propia validez como teoría, pero esto no es tan relevante para este ensayo). En cambio, responder afirmativamente le permite a las TH defender un tipo de teorías metafísicas que serían válidas y suficientes solo apoyándose en las fuentes directas de información que nos da la experiencia.

Las TNH responden a la pregunta de la validez afirmativamente. Nuevamente, lo “posible” hace referencia a “la posibilidad con mantención de la validez”. Cada teoría que caiga en esta clasificación diferirá en exactamente qué inobservables postula y cuáles postulaciones considera válidas, pero comparten su aprobación de este método para explicar el mundo. En el caso de estas teorías, responden negativamente a la pregunta por la suficiencia y, de esta forma, se ven obligadas a justificar por qué una teoría como las TH, que pretende limitarse a la experiencia sin la postulación que vaya más allá de esta como los inobservables, no nos alcanza para explicar satisfactoriamente el mundo.

Todo lo anterior ha servido para mostrar dónde se ubica el desacuerdo, ya que incluso cuando hablamos de “posibilidad” nos encontramos refiriendo a la validez y suficiencia de nuestras teorías metafísicas, nos encontramos con preguntas con una fuerte carga de normatividad con respecto a nuestras metodologías de teorización con respecto al mundo. Al preguntarnos por la forma “más correcta.” “la mejor”, no podemos seguir adelante evadiendo la pregunta “mejor bajo qué criterios” “mejor para qué objetivo”. Esto introduce una parte esencial de cómo se

desenvuelve este desacuerdo; me refiero a las posturas y orientaciones valóricas que guían nuestra decisión entre una respuesta u otra. Apoyándome en lo anterior, pretendo plantear una interpretación del choque entre TH y TNH, entendiéndolo como un desacuerdo de evaluaciones normativas cuyos cimientos son las preguntas sobre lo que consideramos valioso y para qué fines al hacer metafísica.

#### **4. Indagando en las razones del desacuerdo: ¿Mejor bajo qué criterio?**

La respuesta más general a nuestra pregunta sería: basado en aquello que valoro en una teoría metafísica. Esto puede ser altamente variable entre distintas personas, y me parece útil ilustrar el tipo de criterios que considero que estarían en juego al considerar mejor una u otra postura metafísica, algunos ejemplos son los siguientes:

1. Mejor según como presupongo que es el mundo: esto se refiere a que la preferencia por una TH o una TNH está determinada por cómo estas teorías coinciden o no con la preconcepción que tengo acerca de la realidad; estas preconcepciones están influenciadas por las experiencias (cotidianas, socioculturales y con respecto a diversas fuentes de conocimiento de la época) que ha tenido la persona. Por ejemplo, alguien que prefiera una TNH puede hacerlo debido a que cree que el mundo está ordenado de una sola forma definida; esto lo lleva a preferir una teoría que él cree que apunta a explicar en qué se fundamenta el orden del mundo, en vez de una TH que se intenta alejar de los postulados de inobservables necesarios para realizar esta tarea.
2. Mejor basado en las consecuencias prácticas que implica adoptar una u otra postura metafísica: esto se refiere a que la preferencia por una TH o una TNH se produce porque una de estas posturas lleva a consecuencias prácticas que yo valoro positivamente, como lo podrían ser aquellas formas de actuar en la disciplina que se consideran ligadas a quienes toman una postura metafísica de uno u otro tipo (también podría ser al revés si es que se quiere principalmente evitar comportamientos o consecuencias prácticas asociadas a posturas contrincantes). Por ejemplo, alguien puede preferir una TH debido a que cree que al no delimitar una explicación determinada del funcionamiento de la realidad, se favorece la

apertura a diversas teorías y esto él o ella lo evalúa positivamente para la disciplina.

3. Mejor basado en cuáles son los objetivos de la metafísica: es importante considerar cómo dependiendo de qué objetivos le demos a la metafísica (en este caso específico, especialmente su relación con las prácticas científicas) variará nuestra evaluación acerca de las teorías que aportan más o menos a estos objetivos. Esto es verdad incluso si tuviésemos un acuerdo mínimo con respecto a entender la metafísica como una empresa cognoscitiva (que tiene como objetivo, o al menos como uno de los principales, el adquirir conocimiento válido acerca del mundo).

Estos son tan solo tres ejemplos que muestran de dónde podrían provenir las justificaciones de lo que consideramos formas válidas y suficientes de establecer teorías metafísicas respecto a nuestra experiencia del mundo. Con todo, y aunque creo que resultan ilustrativos al resaltar aquellos factores que pueden ser relevantes para la discusión evaluativa entre las posturas TH y TNH en casos particulares, ni de cerca abarcan la cantidad de criterios que definen nuestra evaluación acerca de “la forma correcta o mejor de establecer una teoría metafísica”. Incluso en mi intento de reducir la problemática al caso específico de la decisión de postular inobservables o no hacerlo, hasta ahora no he establecido una idea general acerca de qué depende nuestra decisión.

Creo que esto se debe a que incluso este tipo de decisiones muy específicas en la confección de una teoría metafísica hallan su justificación en un sistema interconectado de creencias respecto a la metafísica misma y lo que es deseable en ella. Tal como mencionábamos antes, se tratan de preguntas normativas respecto a la tarea filosófica de la metafísica y cualquier concepción normativa de este tipo, es decir, respecto a una disciplina cognoscitiva, se encuentra fundamentado en cada uno de los sistemas de creencias que sostienen los seres humanos que las practican y que son aquellos que entablan estas discusiones entre TH y TNH. Este es un tema extremadamente amplio, así que intentaré ser breve en presentar una posible perspectiva al respecto.

## 5. La perspectiva meta-filosófica que estoy adoptando

### Filosofía, punto de vista y racionalidad

Nicolas Rescher, un filósofo principalmente asociado a la corriente pragmatista, en su texto *The Strife of Systems: An Essay on the Grounds and Implications of Philosophical Diversity* intenta responder a preguntas similares a las que antes mencioné, específicamente respecto a cuál es la naturaleza de la disciplina y, en concordancia con esto, cuáles son los medios por los cuales se justifica la adopción de una u otra postura filosófica. En resumen, se entiende la filosofía como una empresa cognoscitiva en el sentido básico que antes mencioné, pero además se le atribuye la labor de, tomando la diversidad de experiencias del mundo con las que nos encontramos en nuestro paso por este, elaborar una sistematización coherente y justificada de nuestra perspectiva del mundo. Se hace especial hincapié en cómo nuestra experiencia nos lleva a situaciones de disonancia cognitiva, en las cuales tenemos razones para sostener varias creencias, pareciéndonos cada una de estas verosímiles por su cuenta, pero que al mismo tiempo nos parecen insostenibles en conjunto.

La filosofía se encarga de establecer cuáles serían estas situaciones aporéticas fundamentales con respecto a nuestra experiencia del mundo, es decir, los problemas filosóficos de los que esta se encargaría, y luego también se encarga de proponer diversas soluciones para solventar estas contradicciones y crear una perspectiva sistemática y coherente; es decir, establecer posturas filosóficas entre las cuales se dan los desacuerdos. Una de las principales ideas del texto hace referencia a que cuando nos planteamos frente a un conjunto aporético (los conjuntos de creencias antes mencionados) e incluso si estamos de acuerdo en que efectivamente se trata de una contradicción que se debe resolver, no existe un punto de vista neutral desde el cual podamos establecer la mejor forma de restablecer la coherencia. Existirán siempre muchas vías, cada una de ellas implicará ciertos costos y beneficios, y cada uno de estos aspectos valorados positiva o negativamente serán evaluados y priorizados de forma distinta dependiendo del punto de vista desde el cual nos encontremos. No existiría una solución que fuese la absolutamente perfecta desde todo punto de vista o según cada aspecto que consideremos valioso en nuestras evaluaciones.

Ahora es importante precisar de qué están compuestos estos puntos de vista. Lo planteado en este párrafo son posturas más que buscan es-

clarecer el argumento anterior inspirándome en las posturas de Rescher. Las fuentes desde las cuales se forma nuestro punto de vista pueden ser muy variadas: experiencias rutinarias del día a día, testimonios de otras personas, formulaciones teóricas a las que nos expongamos, creencias religiosas, tradiciones culturales, entre muchas otras. Por ello, no pretendo hacer una categorización de las fuentes de nuestro punto de vista, pero sí una que se haga cargo de los aspectos constitutivos de nuestro punto de vista, partes distinguibles entre sí que conforman el todo que es nuestro punto de vista y que se desarrollan por medio de las fuentes que antes mencionaba.

Creo que primero podemos establecer el aspecto teórico de nuestro punto de vista, el cual hace referencia a la información con la que contamos hasta un momento dado, aquello que consideramos hechos y datos y, por lo mismo, aquello que utilizaremos como evidencia en nuestros intentos de justificar nuestra perspectiva. Este aspecto es relativo al punto de vista por el hecho evidente de que ninguno de nosotros contamos con algo así como una capacidad de absorber la “información completa” relevante para un problema o situación, y también debido a que es falso que todos nosotros en todo momento contamos con una información idéntica a la que tienen los demás. Esto resulta relevante debido a que nuestras creencias y las justificaciones que formemos de ellas estarán en estrecha relación con la información que tengamos a nuestra disposición. Por otro lado, también hay un aspecto evaluativo de nuestro punto de vista; esto se debe a que atribuimos diversos valores o disvalores a diversos aspectos del mundo, priorizamos unos por sobre otros en nuestros juicios y tomamos decisiones basándonos en estas evaluaciones, siempre desde una perspectiva específica acerca de qué es valioso en cada situación y qué fines deberíamos perseguir. De nuevo, esto resulta sumamente relevante para lo que analizábamos en secciones anteriores, ya que cuando se trata de preguntas sobre qué objetivos debería perseguir una empresa cognoscitiva o cuáles son las herramientas o métodos válidos para el desempeño de esta, necesitamos hacer una evaluación asignando valor a los distintos aspectos en juego. Finalmente, creo que también podríamos separar un aspecto práctico de nuestro punto de vista, el cual tiene que ver con el tipo de acciones que consideramos adecuadas como una decisión óptima respecto a situaciones particulares, sin embargo, este último aspecto no resultará tan relevante para este trabajo, al menos no por su cuenta.

Lo anterior congenia con la idea de racionalidad que Rescher defiende (Rescher 1988), la cual juega un rol en su definición de la filosofía,

ya que se la entiende como una forma óptima de utilizar nuestra capacidad de razonar, lo cual se entiende como la búsqueda inteligente de los mejores medios para perseguir los fines correctos. Tan solo en esta formulación nos encontramos con los tres aspectos antes mencionados: la búsqueda inteligente hace referencia al aspecto teorético, los mejores medios a lo práctico y los fines correctos a lo evaluativo. Por lo mismo, la racionalidad se divide en tres tipos de esta misma, que funcionan como un sistema interconectado entre sí y que, por lo mismo, se necesitan en conjunto para funcionar óptimamente: la racionalidad teorética, la racionalidad práctica y la racionalidad evaluativa. Sin contar con el espacio para profundizar tanto en el detalle de esta postura en el presente texto, vale la pena esclarecer el término optimización ya que se caracteriza por requerir una operación de costo/beneficio, en la cual, sin tener a nuestro alcance un punto de vista teorético, evaluativo ni práctico absoluto y perfecto, tenemos que hacer lo mejor dentro de lo que nos es posible, lo cual generalmente implica compatibilizar múltiples intereses y aceptar el no alcanzar un ideal perfecto en nuestra teorización, evaluación y práctica.

Volviendo a *The Strife of Systems*, ahora podemos introducir la propuesta que considero más relevante mencionar para este trabajo, ya que, según Rescher, cuando nos encontramos frente a estos conjuntos aporéticos, además de nuestro interés por hacer filosofía y restablecer la consistencia entre nuestras creencias acerca del mundo en un todo sistemático y coherente, nos vemos obligados a decidir entre las múltiples formas en que esto es posible, y podemos hacerlo desde el único lugar que tenemos la posibilidad realizar esta elección y justificarla: nuestro punto de vista. Y se trata de una decisión que, entre otros aspectos, requerirá de la evaluación y consiguiente atribución de valor para establecer cuál forma de resolver el problema filosófico nos parece más adecuada, válida y suficiente en comparación con el resto. Según el autor, estas evaluaciones las hacemos con un sustento que él denomina valores cognoscitivos, la formación de estos se daría de forma tan variada como lo que mencionábamos acerca de nuestro punto de vista y las fuentes según las cuales se desarrollaba. De hecho, podríamos inicialmente caracterizar a nuestros valores cognoscitivos como nuestro punto de vista evaluativo, pero esto requiere de esclarecerse.

## Valores cognoscitivos: un concepto que hay que clarificar

En esta sección propondré una forma de entender el concepto de valor cognoscitivo, la cual considero que mantiene ciertos lineamientos de la idea de Rescher, pero que no se encuentra expresado claramente de esta forma en *The Strife of Systems*. Creo que este concepto se entiende mejor describiendo dos aspectos del tipo de evaluaciones que hacemos para decidir entre soluciones a problemas filosóficos y cómo el concepto de valores cognoscitivos intenta congeniar ambas ideas: tendencias valorativas y carácter normativo de nuestras valoraciones.

1. Tendencias valorativas: Tomo como un presupuesto algo mencionado anteriormente, que producto de una variedad de experiencias que tenemos los seres humanos a lo largo de nuestra vida, no solo desarrollamos un punto de vista respecto a los hechos del mundo, sino que también uno respecto a lo que consideramos valioso, importante, deseable, etc. Podríamos, en un primer momento, suponer que la formación de este punto de vista se realiza solo por medio de las instancias en que establecemos juicios explícitos, ya sea en nuestro pensamiento o expresándolo hacia otros, respecto a lo que le atribuimos un valor específico o lo que consideramos prioritario en una situación específica. Sin embargo, considero que esto sería errado debido a que habría múltiples situaciones que requerirían de una evaluación y en la que no encontramos un juicio explícito al respecto. Pero considero que dar ejemplos de ese estilo no sería un argumento tan fuerte como el siguiente: que debido a la misma estructura según la cual se encadenan las justificaciones de nuestras valoraciones, se vuelve poco verosímil la propuesta anterior. Me refiero a cómo si juzgamos explícitamente el valor de algo y nos preguntamos por la justificación de esto, a veces lo justificamos en su conexión con otras cosas que valoramos y así sucesivamente. Entonces, si nuestro punto de vista evaluativo solo se constituyese por los juicios explícitos, no se podría incluir ninguna de las valoraciones que supuestamente justificarían nuestra evaluación porque todos estos están implícitos.
2. Carácter normativo de nuestras valoraciones: Por otro lado, si bien no todas las valoraciones que hacemos en nuestra vida en general tienen por qué pretender aplicar más allá de nuestras preferencias personales, cuando nos referimos a la filosofía como



empresa cognoscitiva, que no solo implica el trabajo de investigación en conjunto con otras personas que también tienen el interés de conocer el mundo, sino que está estrechamente relacionada a las exigencias de nuestra racionalidad, que nos pide tener una visión coherente del mundo, tener buenas razones para nuestras creencias, elegir los fines correctos que perseguir y actuar de la forma adecuada en la persecución de estos, nos vemos obligados a alejarnos de la idea de que la filosofía se tratase de meramente expresar nuestras preferencias personales. Lo anterior, creo que se debe al carácter normativo de la racionalidad, como en la búsqueda de respuestas a nuestras preguntas más profundas respecto al mundo no podemos sino hacerlo bajo estándares que no sean ajenos a esta, como lo serían las predilecciones personales que no necesitarían de justificación más allá del hecho de mi gusto por ellas. Al menos con esta forma de concebir la filosofía y lo que nos interesa conseguir en esta, se necesita de una discusión y justificación de nuestras valoraciones, y si nos negamos a apelar a estándares absolutos que son verdaderos para todos por igual, debemos apoyarnos en la construcción de estándares intersubjetivos acerca de los fines que deberíamos perseguir, aquello que en consecuencia deberíamos valorar y con cuánta prioridad valoramos cada cosa en cada situación (en la práctica filosófica).

Vale la pena resaltar cómo en este trabajo se está haciendo esto mismo, aun si el aporte es pequeño y limitado con respecto a cómo deberíamos concebir la filosofía, especialmente acerca de los intereses que esta persigue y cuáles son las formas valiosas de realizar nuestros juicios y prácticas filosóficas. Sin tener la expectativa de que mis postulados correspondan a una verdad absoluta, pero buscando hacer la mejor reconstrucción posible desde mi punto de vista de un argumento normativo sobre la filosofía misma. Aun así, pretendo que mi discurso haya seguido estándares de justificación que van más allá de solo mis intereses personales, si bien es imposible realizar esto sin contar ya con tendencias valorativas acerca de estos temas, un aspecto sumamente personal de mi punto de vista.

Todo esto para tan solo decir que en la misma configuración de lo que serían los valores cognoscitivos están implicados tanto un aspecto sumamente personal y dependiente de nuestro contexto y subjetividad, como son las tendencias valorativas, y una necesidad de discutir racionalmente respecto a lo que valoramos, justificar mis posturas filosóficas más allá de solo lo que me ocurre personalmente. Ahora podemos volver

a la discusión acerca de las TH y TNH y plantear una posible interpretación de los valores cognoscitivos que cada una estaría defendiendo, aspectos que cada teoría tendería a valorar en general y que requieren de tener un carácter normativo como el descrito anteriormente para ser defendibles como una postura filosófica acerca de si sería válida o no la postulación de inobservables en metafísica.

## **6. De vuelta a las TH y TNH: ¿Qué lineamientos podemos identificar en cada una?**

En este apartado no creo poder dar una respuesta que abarque todos los aspectos en juego en cada postura, pero sí al menos ofrecer ideas centrales sobre cada una. En el caso de las TH, parecen estar principalmente guiadas por los siguientes puntos:

- a) Humildad acerca de nuestras capacidades cognoscitivas: frente a la tarea titánica de tratar de comprender la constitución fundamental de nuestra realidad y teniendo en cuenta la situación en la que nos encontramos los seres humanos, con medios parciales y no absolutos de aprehender el mundo, se valora el adoptar una postura humilde que reconozca nuestras limitaciones en lo que podemos llegar a conocer y que ponga esto en práctica con respecto a los postulados metafísicos, pues se preferiría mantener una perspectiva cuidadosa en vez de decir más allá de lo que realmente somos capaces y engañarnos con respecto a lo que podemos saber del mundo.
- b) Rechazo a posturas totalizantes o especulativas: esto se entiende perfectamente con lo dicho en a), porque son justamente estas posturas las que postulan más allá de lo que se prefiere en las TH; de todos modos, me gustaría destacar cómo en este aspecto resalta el desacuerdo en el ámbito de las conductas y tendencias de quienes postulan este tipo de teorías. Parece plantearse una asociación entre este tipo de posturas y comportamientos que justamente irían en contra de la humildad antes mencionada y resaltarían por algún grado de egocentrismo y tozudez en su forma de discutir sobre metafísica. Creo que sin tener que pronunciarme sobre la realidad de estas asunciones, puedo sostener que estas tienen efecto en quienes se oponen a este tipo de posturas y en cómo guían sus juicios evaluativos.

- c) Pluralidad de opciones: este punto se relaciona con la falta de respuestas metafísicas cerradas y abarcativas de las TH acerca de cómo estaría configurado y constituido nuestro mundo, lo que las dejaría mejor dispuestas a recibir y convivir con una pluralidad de perspectivas que choquen en postura sobre estas temáticas que las TH se niegan a dar una solución determinada. Nuevamente, esto puede verse como una elección dada por los efectos prácticos que tiene en la discusión metafísica y cómo se podría valorar que está en la práctica incorpore una pluralidad de perspectivas, pero también podría tener que ver con la actitud y comportamientos asociados a quien tiene ideas más incluyentes y de aceptación hacia distintas posturas y cómo este comportamiento se considera beneficioso para la disciplina.

En el caso de las TNH, creo que podemos mostrar los siguientes puntos que juegan un rol a la hora de evaluar posturas filosóficas y terminar decantándose por una postura de este estilo:

- d) Espíritu de búsqueda explicativa: en el tipo de posturas que se defienden desde las TNH, creo que se muestra una valoración positiva hacia la búsqueda de explicaciones que sean lo más completas y con estructuras coherentes que nos permitan comprender las bases de nuestra realidad. Esto lleva a que vean como algo preferible el mantener esta búsqueda de explicaciones y fundamentos de aquello que experimentamos, incluso si eso implica afrontarse a la complejidad y a la posibilidad de errar que implican investigaciones que otras posturas metafísicas, como las TH, consideran imposibles o inválidas.
- e) Sistemática y completitud: relacionado con lo anterior, también suele haber una valoración positiva de una perspectiva que se preocupe de la sistemática de sus posturas y que pueda dar cuenta de cómo estas forman un todo coherente y que no deja “cabos sueltos”. Y a su vez, evalúan en base a cómo sus contrincentes, quienes adoptan posturas de las TH, tienden a descuidar este aspecto sistemático y coherente, no solo desde un aspecto teórico, sino que también en sus actitudes y comportamientos frente a las discusiones metafísicas. Al igual que en el punto b), tan solo mantengo la existencia de estas asunciones y cómo afectan a nuestras formas de evaluar. Esto nos plantea en un punto de vista valorativo distinto, aun si no del todo incompatible al del

pluralismo de las TH. Lo que se valore de ciertas posturas contrincantes y en qué grado se prefieran unos factores sobre otros, sí es claramente distinto.

## 7. Consideraciones finales acerca de lo anterior

Luego de haber planteado estos posibles lineamientos valóricos que defenderían las TH y TNH, que nos permiten tener una comprensión mayor de las motivaciones que guían a cada postura a adherir a sus posturas filosóficas, como en el caso de la postulación de inobservables, podríamos simplemente seleccionar la colección de lineamientos que nos parezcan intuitivamente más adecuados, enunciar que adherimos a ellos y quedar conformes y satisfechos. Sin embargo, considero que esto sería insuficiente e inadecuado si adscribimos a la concepción antes mencionada acerca de cuál es la tarea de la filosofía. A continuación, entregaré los argumentos por lo que creo esto.

Primero, tal como indicábamos al analizar el componente normativo de los valores cognoscitivos que guían nuestra decisión al tomar una u otra postura filosófica, no podemos quedarnos conformes con que nuestra adscripción a un punto de vista valorativo sea una cosa de mero agrado personal, debemos desarrollar justificaciones que tienen un sustrato personal inevitable, pero que apuntan a una discusión racional intersubjetiva y cuyos argumentos están diseñados para lograr más que solo enunciar mis preferencias personales. Hacer filosofía restringiéndonos a esto último sería negarse a intentar responder las preguntas más complejas de lo que anteriormente llamamos mundo, esto solo es defendible si desprendemos a la filosofía completamente de su carácter como empresa cognoscitiva. La metafísica es una de las áreas de la filosofía en que es especialmente difícil sostener lo anterior, ya que no podemos quitarle a la metafísica la labor de la búsqueda de conocimiento acerca de los aspectos más fundamentales del mundo sin a la vez estar perdiendo el sentido de la disciplina entera. Podemos tener desacuerdo respecto a muchos otros puntos sobre cómo es más correcto hacer investigación metafísica, pero no quitarle a esta el carácter investigativo y de búsqueda de conocimiento.

Segundo, por la misma necesidad de justificación y por cómo funciona nuestro punto de vista desde el que justificamos, no podemos hacer esto por la mera adscripción a lo que otros dicen (al menos no literalmente solo absorber las posturas de otros al respecto). Cualquiera de las preguntas planteadas a lo largo de este trabajo, como si es válido

postular inobservables como una metodología para investigar metafísicamente el mundo, si la experiencia nos entrega suficiente material como para entender satisfactoriamente cómo funciona el mundo, si debemos perseguir la completitud de nuestras teorías, si en cambio es preferible no arriesgarse a caer en una ilusión de que conocemos más de lo que realmente nos es posible, si debemos aceptar otras teorías en búsqueda de diversidad o deberíamos criticarlas en búsqueda de exactitud y tal vez la más importante desde nuestro punto de vista: dónde se encuentra el punto de equilibrio entre todos estos factores en juego que nos parezca la forma más razonable de proceder al teorizar en metafísica. Son todas preguntas que se responden desde el punto de vista de uno con todas las peculiaridades de mis tendencias valóricas y la forma en la que justificar normativamente este tipo de posturas. Evidentemente, podemos establecer acuerdos con otros e incluso esos acuerdos a veces son tan fuertes que generar una postura filosófica común a un grupo, pero eso no significa que cada investigador pueda adoptar literalmente la postura del conjunto de ellos, siempre la justificación de nuestras posiciones las hacemos desde nuestro punto de vista.

Y tercero, considero que es imposible adoptar los lineamientos de las TH o TNH sin mayor matización, ya que entre estos lineamientos valóricos nos encontramos con valoraciones de distinto tipo, al menos en tanto son valoraciones que aplican a tipos de objetos distintos, siendo el ejemplo más evidente el que algunos se refieren al comportamiento de los/las colegas investigadores/as y otros a la formulación de nuestras teorías. Esto nos obliga a tener que emitir otra evaluación desde nuestro punto de vista evaluativo respecto a cuáles de estos aspectos priorizamos como más relevantes al tomar una postura u otra, por ejemplo, hasta qué punto debería ser relevante el que las actitudes de un colega puedan ser cerradas en la discusión con otras posturas en comparación con que su teoría tenga una sistematicidad y capacidad explicativa muy valorables. Sin siquiera empezar a preguntarnos por la inmensa variabilidad de posibles respuestas con la que nos encontraríamos al decidir cómo balancear todos estos aspectos y adoptar una postura filosófica, ya resulta claro que no es posible simplificar esta decisión a la mera aceptación de los lineamientos de las TH o TNH.

Todo lo anterior tiene como propósito llevarnos a concluir lo siguiente. Incluso cuando nos encontramos con una pregunta tan específica como ¿Debería postular inobservables al hacer metafísica? Y tenemos las respuestas estereotípicas y simplificadas de la postura afirmativa (TNH) y la negativa (TH), no podemos esperar responder sin acudir a un punto

de vista valorativo y consideraciones normativas. E incluso si tenemos la versión simplificada de los lineamientos valóricos que guiarían a las TNH y las TH en su comprensión de la tarea de la metafísica y cuál sería la mejor forma de proceder en esta empresa cognoscitiva, no hay una respuesta fácil al problema; es decir, una que no requiera una contundente justificación evaluativa acerca de cómo deberíamos priorizar y valorar diversos aspectos al hacer metafísica. Espero con esto haber mostrado la complejidad que es necesaria atender al hacernos esta pregunta y, a su vez, el tipo de errores que fácilmente podemos cometer si no somos conscientes de qué tipo de discusión es la que estamos teniendo y qué justificaciones necesitamos para defender nuestras posturas al respecto. Al responder a las preguntas sobre los inobservables en metafísica, lo que se pone sobre la mesa son nuestras tendencias valorativas y valoraciones normativas, es decir, nuestros valores cognoscitivos, y no lograremos dar respuestas sólidas a este problema si no atendemos al sustento que justifica nuestra visión más general sobre la metafísica como empresa cognoscitiva.

## Referencias

Armstrong, David. (1983). *What is a law of nature?* Cambridge University Press, New York. DOI: 10.2307/2185228.

Lewis, David. (1986). *Philosophical Papers, Volume 1*. Oxford University Press. DOI: 10.1093/0195032047.001.0001

Lewis. (1994). *Humean Supervenience Debugged*. Oxford University Press. <https://www.jstor.org/stable/i339346>

Rescher, Nicholas. (1985). *The Strife of Systems: An Essay on the Grounds and Implications of Philosophical Diversity*. University of Pittsburgh Press. DOI: 10.2307/2107718

Rescher. (1988). *Rationality: a philosophical inquiry into the nature and the rationale of reason*. New York: Oxford University Press. <https://searchworks.stanford.edu/view/1323447>

Soto, Cristian. (2024). *Leyes de la naturaleza: Historia, Filosofía y Ciencias*. Editorial Tecnos. <https://www.tecnos.es/libro/filosofia-y-ensayo/leyes-de-la-naturaleza-cristian-soto-9788430990375/>

# La propuesta de Hempel para un análisis de los juicios morales a partir de los métodos de contrastación y validación científicos

Rodrigo López-Orellana<sup>a</sup>

## Resumen

El presente trabajo tiene por objetivo recordar y explorar críticamente, de manera muy breve, la propuesta del análisis de la valoración moral de C. G. Hempel en 1965. Hempel incluyó el análisis de los juicios morales en la perspectiva objetiva y causal de la ciencia. Su intención era lograr que la evaluación moral pudiera resultar también en un sub-modelo probabilístico-inductivo, lo que le lleva a considerar —a su vez— que el conocimiento y el método científico presuponen la evaluación moral.

**Palabras clave:** valores, moral, ciencia, explicación, causalidad, objetividad, Hempel.

## Abstract

The aim of this paper is to recall and critically explore, in a very brief way, C. G. Hempel's proposal of the analysis of moral judgement in 1965. Hempel included the analysis of moral judgements in the objective and causal perspective of science. His intention was to ensure that moral evaluation could also result in a probabilistic-inductive

---

<sup>a</sup>Instituto de Filosofía, Universidad de Valparaíso, Chile. Instituto de Estudios de la Ciencia y la Tecnología (ECYT), Universidad de Salamanca, España. <https://orcid.org/0000-0002-3576-0136>.

Contacto: [rodrigo.lopez@uv.cl](mailto:rodrigo.lopez@uv.cl), [rodrigo.lopez@usal.es](mailto:rodrigo.lopez@usal.es)

sub-model, which leads him to consider —in turn— that scientific knowledge and method presuppose moral evaluation.

**Keywords:** values, moral, science, explanation, causation, objectivity, Hempel.

## 1. Introducción

Carl G. Hempel, en su *Scientific Explanation* de 1965 —específicamente en el ensayo recopilado *Science and Human Values*, escrito originalmente en 1960—, esbozó un tratamiento acerca de los juicios morales de valor desde una teoría de la explicación por subsunción (*Subsumption Theory of Explanation*, o también *Covering Law Model*). Para Hempel, la investigación científica y sus métodos de validación o justificación de la ciencia empírica tenían un cierto paralelismo con el caso de la valoración y decisión moral, ya que creía que el método científico podía desempeñar un papel importante en la clarificación y solución de problemas en el ámbito de la acción moral. Estaba convencido de que la investigación científica podía entregarnos la información fáctica necesaria para la resolución de este tipo de problemas. Todo esto se enmarca en su intento de demostrar que su perspectiva de la explicación científica tenía también validez para las *non-physical sciences*: que el análisis causal era correcto también para los *purposive behaviors*. Al respecto de esta peculiar propuesta de Hempel, de una sobre extensión del enfoque de la “Concepción heredada”, G. H. von Wright (1971) decía:

[...] parece casi una ironía del destino el que la formulación más completa y lúcida de la teoría positivista sobre la explicación llegara a establecerse a propósito de la materia para la que, obviamente, la teoría se halla peor dispuesta, a saber, la historia. Pero es probable que sea principalmente por esa misma razón que el artículo de Hempel haya provocado una inmensa cantidad de discusión y controversia (von Wright 1971, p. 10-11).

Exploraremos, a continuación, esta propuesta de Hempel.

## 2. Análisis de los juicios morales de valor desde una perspectiva objetiva y causal

Según Hempel (1962b, 1965), los métodos de validación científicos son aquellos que le permiten a la ciencia examinar una hipótesis nueva



y a la vez determinar si debe ser aceptada o rechazada. Generalmente, los métodos de testeo y validación involucrados en la determinación de aceptabilidad de una hipótesis son métodos indirectos, ya que, en los casos de algunas de las hipótesis más importantes de las ciencias, la observación directa de hechos no siempre basta para decidir si éstas deben ser aceptadas como verdaderas o rechazadas como falsas.

Desde este enfoque, si se supone que la hipótesis principal investigada es verdadera, debe inferirse de algún modo otros enunciados que describan fenómenos directamente observables, que deben manifestarse en circunstancias especificables previamente. En efecto, los enunciados inferidos se testean directamente, en otras palabras, se verifica si tales fenómenos especificados ocurren realmente. Según los resultados, la hipótesis propuesta se rechaza o se acepta y a medida que se verifica un número cada vez mayor de éstos, la hipótesis comporta un grado mayor de aceptabilidad. Para Hempel, se debe tener presente que nunca es posible probar de manera completa y definitiva una hipótesis científica, siempre está la posibilidad teórica de que se descubran nuevos elementos de juicio que contradigan o produzcan conflicto con aquellos enunciados observacionales, conduciéndonos así —ineludiblemente— al rechazo de la hipótesis de la cual son inferidos.

...  
Un ejemplo clásico que confirma este ejercicio de método en ciencias es el caso de *la Tierra tiene forma esférica* y su comprobación indirecta a través de la experimentación de la curvatura de la Tierra mediante la observación de otros fenómenos, como el alejamiento de los barcos en el horizonte.

Lo que le interesaba a Hempel era ver si este método podía ser usado para establecer juicios morales de valor. ¿Es posible que este método sirva para establecer juicios según los cuales una alternativa de acción sea buena, correcta o más apropiada frente a otras alternativas de causar tal acción? Para Hempel, existe un tipo de esta clase de enunciados que cae dentro de la esfera de la investigación científica, y que representa la aplicabilidad del método: el juicio de valor relativo o instrumental. Este es un juicio de forma condicional, a saber:

[...] cierto tipo de acción,  $M$ , es bueno (o es mejor que otro tipo alternativo  $M_1$ ) si se quiere lograr un objetivo  $O$  determinado. Pero decir esto equivale a afirmar, o bien que, en las circunstancias dadas, el curso de acción  $M$  conducirá definitivamente (o probablemente) al logro de  $O$ , o bien que

no seguir el curso de acción  $M$  conducirá definitivamente (o probablemente) al no logro de  $O$  (Hempel 1965, p. 84).

Esos juicios se reformularán en enunciados que expresen un tipo de relación ya sea universal o probabilística entre medios y fines, excluyendo la carga moral que comportan, esto quiere decir, no incluyendo términos del discurso moral (“bueno”, “malo”, etc.): se afirmará o que 1.  $M$  es un medio suficiente para alcanzar un fin u objetivo, o que 2.  $M$  es un medio necesario para alcanzar dicho fin u objetivo (ya sea definida o probablemente).

Por ejemplo, podríamos tener los siguientes enunciados como resultado de alguna investigación que hayamos hecho:

1. El príncipe que actúa con rapacidad para con los bienes ajenos y actúa con atropello de las mujeres de sus súbditos es hallado odioso, digno de menosprecio, pusilánime e irresoluto, sin cumplimiento de su deber, convirtiéndose en un mal gobernador.
2. El príncipe que actúa con dignidad respetando los bienes y mujeres de sus súbditos no será hallado odioso ni despreciable, los hará felices y habrá cumplido con su deber, y no quedará más que combatir las ambiciones de unos pocos.

Según Hempel, si tales enunciados se amplían adecuadamente, como vimos más arriba, estableciendo que se trata de enunciados de tipo condicional, es fácil ver cómo podemos desarrollar una explicación causal donde este tipo de enunciados resultan ser afirmaciones empíricas capaces de ser sometidas al test empírico:

Si el príncipe quiere ser respetado y apreciado, y resultar ser un buen gobernador, entonces es mejor que respete los bienes y mujeres de sus súbditos y no actúe con rapacidad y atropello.

Esta afirmación puede testearse empíricamente si revisamos los ejemplos que nos da la historia, y podríamos desarrollar así un juicio instrumental de valor que nos sirva para conducir la conducta futura, si es que queremos llegar a ser un buen príncipe o gobernador. Pero existe otro tipo más de juicios de valor que no resulta tan fácil aplicarle este método.

El problema con la aplicación de este método se produce ante los juicios categóricos o absolutos de valor, un tipo de juicio según el cual

cierto estado de cosas (propuesto ya sea como fin u objetivo) es bueno o mejor que otra alternativa especificada. ¿Cómo podemos someter a un testeo empírico y recibir una confirmación empírica a tal tipo de juicios que establecen una primacía de valor moral a un estado de cosas frente otras alternativas? Aquí aparecen enunciados como *matar es malo* o *robar es malo*. Siguiendo a Hempel (1965, p. 85-96), la primera oración, por ejemplo, implicaría calificar categóricamente la eutanasia como un mal y la aserción no puede ser testeadada directamente por la observación, ya que no propone describir un hecho. No podemos decir de ella que sea verdadera o falsa, sino más bien sirve para expresar un patrón para la evaluación moral, o sirve como norma de conducta.

No comporta un contenido empírico descriptivo, por eso hablamos de ella como norma. Tampoco es posible sustentar un sistema de normas morales en hallazgos de las ciencias empíricas, como por ejemplo deduciendo juicios de valor a partir de los resultados de investigaciones empíricas acerca de las necesidades humanas, presuponiendo que es bueno satisfacer necesidades humanas.

La presuposición es un juicio categórico de valor en sí misma, sería una premisa con un contenido evaluativo absoluto, dificultando inmediatamente alguna prueba de testeo empírica. La explicación implicaría una relación compleja entre el *explanans* y el *explanandum*, ya que la premisa evaluativa tendría que ser un complejo sistema de normas que refiere a reglas específicas acerca del orden y grado preferenciales en los que deben satisfacerse tales necesidades. Existen necesidades diferentes y conflictivas, según cada individuo o grupo, lo que hace difícil una deducción de normas evaluativas a partir de un estudio fáctico de las necesidades. Si bien es cierto que es claro ver esta dificultad de una deducción de este tipo, para Hempel aún es posible establecer esas normas para justificar una decisión o un juicio de valor moral a partir de una visión pragmática acerca de lo que puede servir como fuente de evaluaciones incondicionales (Hempel 1983, p. 85-87). De hecho, los seres humanos obtienen sus valores más básicos de esa fuente (Hempel 1965, p. 87).

Desde ese punto de vista pragmático, Hempel inscribe la valoración moral en el ámbito de la ciencia empírica, cuestión que se explica brevemente a continuación.

### 3. La evaluación moral como sub-modelo probabilístico-inductivo

Todos los individuos, o grupos, adquieren sus valores de la sociedad a la cual pertenecen, de su familia, de su iglesia, de las instituciones sociales en las cuales participan, etc. En general, no son valores obtenidos según el resultado de un examen cuidadoso y crítico de las mejores alternativas que le pueda ofrecer determinada sociedad, o grupo, y es muy raro que tales valores sean puestos en examen. Para Hempel, generalmente las decisiones que tomamos y nuestros actos no se basan en ningún conjunto de normas morales explícitas conscientemente adoptadas:

De hecho, en muchas situaciones decidimos y actuamos irreflexivamente en un sentido aún más fuerte, es decir, sin ningún intento de justificar nuestras decisiones en algún conjunto explícito, conscientemente adoptado, de estándares morales (Hempel 1965, p. 87).

Pero cuando queremos justificar nuestras decisiones, las dificultades a las que se enfrenta el modelo de Hempel son, por ejemplo: ¿cómo podemos establecer normas morales que nos permitan esa justificación si la ciencia no nos provee de juicios categóricos de valor? ¿De dónde obtenemos tal sistema de normas que nos permita validar un sistema de juicios categóricos de valor sin caer en contradicciones o en explicaciones circulares, ni caer en afirmaciones cerradas donde la información fáctica involucrada refiere a un único sistema de valoración moral que sólo valen para dicho sistema, como la Torá, el Corán, la Biblia, etc.? (McIntyre 1998, pp. 85-87, 128-130; Dray 1957).

De este modo, para completar el proceso de justificar una decisión o un juicio moral dados, deben aceptarse ciertos juicios de valor sin justificación ulterior, así como la demostración de un teorema geométrico exige la aceptación sin prueba de ciertas proposiciones como postulados (von Wright 1971, p. 34-37; Hempel 1965, p. 85-89):

La búsqueda de una justificación de todas nuestras evaluaciones pasa por alto esta característica básica de la lógica de la validación y la justificación (Hempel 1965, p. 87-88).

Esta comparación de la valoración moral con la investigación y conocimiento científico en base a una misma característica peculiar, a saber,

que existe una aceptación y adecuación de premisas sin una justificación ulterior, se intentó como solución a partir de cierto paralelismo que hacía posible la aplicabilidad del mismo método en ambos casos. En este punto, debemos advertir que Hempel sólo explicó de manera muy sucinta cómo podría desarrollarse este procedimiento dentro de un modelo de explicación como lo es el modelo de cobertura legal para el caso de los juicios categóricos de valor. Sólo más tarde von Wright (1971) se atrevería a una sistematización de la explicación causal de la acción; por supuesto, resultando en un fracaso absoluto.

No deja de llamar la atención que este intento sistemático de von Wright se llevara a cabo luego de entrada la crisis en filosofía de la concepción heredada de la ciencia, especialmente con las consideraciones acerca del método que realiza Kuhn y su inclusión del método de la comprensión (histórica) en el análisis de las ciencias (Kuhn 1962), que incluye el carácter de subjetividad de los cambios científicos.

Hempel mencionará otra característica común entre la valoración moral y la investigación científica. Si comprendemos el universo desde la concepción que tiene la teoría física actual, específicamente a través de los exitosos resultados de la mecánica cuántica y de la idea de que el universo no es un sistema causal cerrado —determinista—, sino más bien que sus leyes tienen un carácter probabilístico, comprenderemos también que nunca alcanzaremos un estado total del conocimiento. Solo tenemos un conocimiento incompleto de las leyes de la naturaleza y del estado del mundo a la hora de decidir sobre nuestra actuación; tomaremos decisiones siempre a la base de una información incompleta, sólo estamos capacitados para prever las consecuencias de las alternativas con un grado especificable de probabilidad. Para Hempel (1962a), es aquí donde la ciencia nos puede brindar cada vez una mejor información para completar una evaluación para (o sobre) nuestra acción. Y en última instancia, aunque siempre tengamos que evaluar nosotros mismos las probables consecuencias de las diferentes alternativas del caso, siempre podemos contar con (y adoptar adecuadamente) normas evaluativas atinentes a ese caso —aunque no estén determinadas objetivamente por los hechos empíricos—.

Hempel relaciona esto con la teoría de adopción de decisiones en matemáticas, que según reglas de decisión determinan una elección óptima frente a diferentes posibles cursos de acción. Incluso, en la teoría de la computación existen programas llamados de computabilidad evolutiva, donde se analiza varias posibilidades de desarrollo genético, tanto las óptimas como las que provocan mutaciones y deformaciones, el sistema

entrega siempre cuáles serían las mejores opciones de decisión según los datos previos entregados (Hempel 1965, p. 87).

Para Hempel, la evaluación y la decisión presuponen la investigación y el conocimiento científicos. La evaluación moral puede incluirse en el sub-modelo probabilístico-inductivo del modelo general de explicación por cobertura legal (von Wright 1971, p. 118-124). Según Hempel, lo más interesante sería ver si existe la contraparte a esta cuestión de que el conocimiento y el método científico presuponen la evaluación moral, lo que se convertiría en una segunda relación existente entre los métodos de contrastación científica y la valoración moral.

#### **4. El conocimiento y el método científico presuponen la evaluación moral**

Hempel analizará entonces si el corpus del conocimiento científico —ese sistema de enunciados de la ciencia que ésta acepta de acuerdo con sus propias reglas de investigación— presupone juicios de valor. Pero no en el sentido de si es posible que tales enunciados puedan inferir lógicamente esa clase de juicios. Ya señalamos más arriba que es imposible que ese sistema implique lógicamente juicios incondicionales —categóricos— de valor. También podríamos aducir que las razones por las cuales aceptamos ese conjunto de enunciados científicos incluyen ciertos juicios incondicionales de valor, pero ya sabemos que sólo bastan los juicios empíricos para dar razones para aceptar o rechazar una hipótesis; esos juicios empíricos incluyen hallazgos observacionales, teorías y leyes previamente aceptadas que comportan la fuerza explicativa para su aceptación o rechazo. No tiene sentido incluirle esa clase de juicios de valor a la explicación, ya que no contribuyen ni a sustentar ni a rechazar la hipótesis, debido a que carecen de toda atingencia lógica.

Para Hempel, el sistema de enunciados del conocimiento científico tiene suficiente apoyo de los elementos de juicio empírico de los que dispone, es un sistema aceptado de acuerdo con los principios del testeo y de la validación científica, aunque los elementos de juicio observacionales que se tienen en el momento no basten para aceptar de manera concluyente o definitiva una hipótesis científica (Hempel 1942).

El riesgo inductivo que se corre en que una ley científica no comporte toda la generalidad que se le atribuye y que los elementos de juicio posibles a futuro lleven a modificar o abandonar la ley, no hace que los científicos rechacen dicha ley. Hempel señala que la ciencia ha desarrollado para la solución de este problema ciertos tipos de reglas para el

procedimiento de este método, reglas de confirmación y reglas de aceptación (Hempel 1965, p. 73-39). Las primeras especifican qué tipos de elementos de juicio son confirmatorios y cuáles son des-confirmatorios para una hipótesis. Las segundas establecen cuán fuerte tiene que ser el apoyo de una hipótesis en los elementos de juicio empíricos para que sea aceptada dentro de ese corpus del conocimiento, esto plantea grados de aceptabilidad de una hipótesis.

En efecto, podemos observar que estas reglas son reglas de decisión y esto hace que exijan la especificación previa de valoraciones que servirán como normas de adecuación, se asignan valores y disvalores previamente definidos a los diferentes resultados posibles de la aceptación o rechazo. Y es aquí donde podemos decir que el método presupone valoraciones: la justificación de las reglas de aceptación y rechazo exige una referencia a juicios de valor (Hempel 1965, p. 90-93). Una hipótesis aceptada es la base de un curso específico de acción que llevará al éxito o al fracaso de la aplicación práctica de la teoría.

Hempel señalará que los resultados posibles de la decisión, en la aplicación de dichas reglas para la aceptación de una hipótesis dada, pueden dividirse en cuatro tipos:

- (1) se acepta la hipótesis (como presumiblemente verdadera) de acuerdo con la regla y de hecho es verdadera;
  - (2) se rechaza la hipótesis (como presumiblemente falsa) de acuerdo con la regla y de hecho es falsa;
  - (3) se acepta la hipótesis de acuerdo con la regla, pero de hecho es falsa;
  - (4) se rechaza la hipótesis de acuerdo con la regla, y de hecho es verdadera.
- La ciencia aspira a realizar los dos primeros casos: la posibilidad de los otros dos representa el riesgo inductivo que debe correr toda regla de aceptación (Hempel 1965, p. 92).

Para Hempel, aun cuando nos movamos dentro del ámbito de la ciencia pura, donde no buscamos una aplicación práctica a ese sistema de hipótesis confirmadas, todavía existe aquí un propósito general de obtención de un cierto fin que él describe como “el logro de un corpus de información confiable, amplio y teóricamente sistematizado acerca del mundo” (Hempel 1965, p. 93). La ciencia presupone la valoración ya que nos habla acerca del mundo, y nos da herramientas para saber de él y para conducirnos ante él. Según esta idea de Hempel, cualquier sistema de normas de procedimiento que nos posibilite hablar verdaderamente acerca del mundo, como el de la ciencia, debe elaborarse en base a los

objetivos que se quieren alcanzar, y su justificación debe ser relativa a esos objetivos, que presuponga dichos objetivos. Es en este sentido que Hempel habla de ‘presuponer’, bajo el ideal de objetividad científica.

Más aún, según esta perspectiva el método científico puede servirnos para la clarificación y solución de los problemas de valoración y decisión moral. En vista a los óptimos métodos de validación y testeo científicos y a la posibilidad de análisis que nos abre los métodos de decisión, el conocimiento y la investigación científica pueden brindarnos toda la información fáctica necesaria para la aplicación de nuestras normas morales, para la resolución de problemas en este ámbito. Necesitaríamos dicha información para seguir cualquier curso específico de acción según los datos que podamos observar acerca de las posibles situaciones, no importando el sistema de valores que se adopte; para elegir siempre el mejor camino como una decisión racional y responsable.

Como podemos observar, este enfoque tiene la creencia en que el análisis de la información fáctica podría mejorar el sistema de normas, o cambiar a mejor definitivamente algunas de nuestras valoraciones. Para Hempel, hacer uso de este recurso nos permitiría saber:

1. si un objetivo en vista puede ser alcanzado en una situación determinada;
2. si es posible alcanzarlo, por qué medios alternativos y con qué probabilidades;
3. qué efectos colaterales y consecuencias ulteriores puede tener la elección de un medio determinado, aparte de la probabilidad de permitir el logro del fin deseado;
4. si varios objetivos deseados son realizables conjuntamente o si son incompatibles, en el sentido de que la realización de algunos de ellos excluya definida o probablemente la realización de otros (Hempel 1965, p. 93-94).

## 5. Consideraciones finales

En la propuesta de Hempel subyace la creencia en que la información que nos brinda la ciencia puede mejorar nuestro sistema de valores, no en base a una ‘des-confirmación’ de valores —si pudiéramos usar ese término—, obviamente, sino provocando el cambio en la apreciación de tales problemas morales. Hempel estimará que la ciencia puede ayudarnos en la aclaración de problemas de valoración mediante un estudio



objetivo de los factores que influyen en la toma de decisiones de un individuo o de una sociedad y que determinan las conductas, mejorando nuestra visión de mundo y ayudándonos a ver las alternativas que existen a ese sistema de normas que tengamos, dándonos mejores opciones para evitar las malas consecuencias.

No querrá decir con esto que lo que se esté haciendo es ‘validar’ dicho sistema de normas morales, sino que el uso de este método nos permite efectivamente la explicación en este ámbito del conocimiento, y a su vez nos permite objetivamente la aclaración y mejoramiento de ese sistema de normas.

De esto se puede colegir que los métodos científicos señalados nos permiten establecer criterios objetivos para la evaluación y pueden suministrar, si se les aplica adecuadamente, criterios para adoptar normas morales válidas, para determinar las conductas más apropiadas para nuestros asuntos individuales y sociales.

Como sabemos de sobra, este modelo hempeliano de explicación es un modelo típicamente neo-positivista, donde el papel de la ley natural es fundamental: enuncia la concomitancia regular o uniforme de fenómenos. La ley es o una implicancia universal, o una correlación probabilística que especifica la causalidad. El problema que tendrá este modelo de Hempel es que si tiene la capacidad de lograr efectivamente (cabalmente) hacerse cargo de la explicación de la acción y de la intencionalidad involucrada (Kitcher 1989; Fetzer 2000). Por ahora, podríamos señalar brevemente, que los aspectos de la intencionalidad de la conducta los reduce al problema de la causalidad.

El método científico no puede guardar relación o paralelismo con la evaluación y decisión moral, y para explicar la acción no se puede ir en búsqueda de causas, los motivos no son acaecimientos y, por lo tanto, no pueden ser del tipo adecuado para ser causas. Muchas son las críticas que se introducen aquí (Dray 1957; Apel 1967; Fetzer 2000; Ryle 2002), pero por lo pronto hemos recordado aquí cómo Hempel propuso inscribir el problema moral dentro del ámbito científico, donde el método científico de explicación tendría un alcance también para la evaluación y la decisión moral.

Por ahora, resultará interesante cómo el enfoque hempeliano, como uno de los máximos representantes de esa concepción heredada, incluye el problema de los valores morales dentro del tratamiento de una investigación científica, causal, objetiva y empírica.

## Referencias

- Apel, K.-O. (1967). *Analytic Philosophy of Language and the Geisteswissenschaften*. Berlín: Springer-Science+Business Media.
- Dray, W. H. (1957). *Laws and explanation in history*. Oxford: Oxford University Press.
- Fetzer, J. (2000). The Paradoxes of Hempelian Explanation. En J. Fetzer (Ed.), *Science, Explanation, and Rationality: Aspects of the Philosophy of Carl G. Hempel* (pp. 111-137). Oxford: Oxford University Press.
- Fetzer, J. (2014). Carl Hempel. En *The Stanford Encyclopedia of Philosophy*: <http://plato.stanford.edu/entries/hempel/>
- Hempel, C. G. (1942). The Function of General Laws in History. *Journal of Philosophy*, 39(2), pp. 35-48. <https://doi.org/10.2307/2017635>
- Hempel, C. G. (1962a). Explanation in Science and in History. En R. G. Colodney (Ed.), *Frontiers of Science and Philosophy* (pp. 9-33). Pittsburgh: University of Pittsburgh Press.
- Hempel, C. G. (1962b). Rational Action. *Proceedings and Addresses of the American Philosophical Association*, 35 , pp. 5-23. <https://doi.org/10.2307/3129344>
- Hempel, C. G. (1965). *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. London: Collier-Macmillan.
- Hempel, C. G. (1983). Valuation and Objectivity in Science. En R. S. Cohen and L. Laudan (Eds.), *Physics, Philosophy, and Psychoanalysis: Essays in Honor of Adolf Grunbaum* (pp. 73-100). Dordrecht: Kluwer.
- Maquiavelo, N. (1983). *El príncipe*. Madrid: SARPE.
- McIntyre, L. C. (1998). *Laws and Explanation in the Social Sciences*. Boulder: Westview Press.
- Kitcher, P. & Salmon, W. C. (Eds.) (1989). *Scientific Explanation*. Minneapolis: University of Minnesota Press.

Rescher, N. (2005). *Studies in 20th Century Philosophy*. Frankfurt: Ontos Verlag.

Ryle, G. (2002). *The concept of mind*. Chicago: University of Chicago Press.

Salmon, W. C. (1980). Probabilistic Causality. *Pacific Philosophical Quarterly*, 61(1-2) , pp. 50-74. <https://doi.org/10.1111/j.1468-0114.1980.tb00004.x>

von Wright, G. H. (1971). *Explanation and Understanding*. London: Routledge & Kegan Paul.



# Normativismo, naturalismo y lo patológico del delirio

Pablo López-Silva<sup>a</sup> y Claudio Anabalón<sup>bc</sup>

## Resumen

Comúnmente, el delirio es caracterizado como un estado mental patológico, una señal de que algo no anda bien en la mente. Sin embargo, existe un debate en la filosofía de la ciencia respecto de qué es lo que haría exactamente a un delirio patológico, sobre todo, en virtud de que sus expresiones paradigmáticas se observan en condiciones psiquiátricas de alta complejidad tales como la esquizofrenia y la depresión mayor. Este capítulo revisa dos alternativas a este problema, esto es, el normativismo y el naturalismo. Luego de analizar sus principales propuestas y respectivas debilidades, se concluye que ninguna de ellas por sí solo ofrece una alternativa lo suficientemente fuerte. Se concluye que las posiciones que combinan ambos enfoques podrían ser una opción plausible en el marco del debate en cuestión y se analiza como éste asunto podría aplicarse al caso del delirio.

---

<sup>a</sup>Escuela de Psicología, Universidad de Valparaíso, Chile. Hontaneda 2653, Of. 4.2., Valparaíso, Chile.

Instituto de Sistemas Complejos de Valparaíso y Instituto Milenio para la Investigación de Depresión y Personalidad.

The Collaborating Centre for Values-based practice in Health and Social Care, St Catherine's College, Universidad de Oxford.

Contacto: pablo.lopez.silva@gmail.com

<sup>b</sup>Instituto de Filosofía, Universidad de Valparaíso, Chile.

Contacto: claudio.anabalonm@postgrado.uv.cl

<sup>c</sup>Este trabajo fue financiado por el proyecto FONDECYT regular n° 1221058 otorgado por la Agencia Nacional de Investigación y Desarrollo (ANID). PLS agradece el auspicio del proyecto Renewing Phenomenological Psychopathology (Discretionary International Exchange Award / 223452/Z/21/Z) otorgado por la Wellcome Trust, UK a Matthew Broome y Giovanni Stanghellini.

**Palabras Claves:** Delirio, Normativismo, Naturalismo, Patología, Enfermedad

## Abstract

Delusion is commonly characterized as a pathological mental state, a sign that something is wrong with the mind. However, there is an ongoing debate in the philosophy of science regarding what exactly makes a delusion pathological, especially since its paradigmatic expressions are observed in highly complex psychiatric conditions such as schizophrenia and major depression. This chapter reviews two approaches to this problem: normativism and naturalism. After analyzing their main proposals and respective weaknesses, it is concluded that neither approach, on its own, offers a sufficiently robust solution. It is suggested that positions combining both approaches could offer a plausible alternative within the framework of this debate, and the application of this issue to the case of delusion is discussed.

**Keywords:** Delusion, Normativism, Naturalism, Pathology, Disease

## 1. Introducción

Históricamente, el delirio ha sido referido como la marca por excelencia de la locura (Jaspers 1963). Por mucho tiempo, ser delirante era sinónimo de estar loco. En gran medida esta visión se mantiene en la actualidad vinculado al hecho de que las formas más complejas y bizarras de aquello que denominamos delirio tienden a observarse en el marco de alteraciones mentales severas como la esquizofrenia y otros trastornos psiquiátricos (Berrios 1991; Beer 1996; Coltheart, Langdon & McKay 2011; López-Silva et al. 2024). Desde un punto de vista conceptual, el delirio ha sido predominantemente caracterizado como un tipo de creencia anormal en lo que se ha denominado el enfoque doxástico (Bortolotti 2010; 2022; 2023; Bayne 2004; 2005; López, Núñez de Prado-Gordillo & Fernández 2023; López-Silva 2023a). Este enfoque ha servido como fundamento conceptual para los principales modelos neuropsiquiátricos (como el modelo de dos factores y el del procesamiento predictivo, ver López-Silva 2023b) y psicoterapéuticos (como el de la Terapia Metacognitiva y la Terapia Cognitivo Conductual, ver Bayne & Pacherie 2005; López-Silva & McClelland 2023) del delirio en la actualidad.

A la luz de la supuesta naturaleza doxástica del delirio, durante los últimos años una serie de filósofos de la ciencia y la psiquiatría han comenzado a preguntarse qué es lo que realmente hace patológico a un delirio (Bortolotti 2015; 2020; 2022; 2023; Bortolotti & Miyazono 2015; Miyazono 2015; 2019; 2022). Esta pregunta es frecuentemente

formulada cuando tales autores comparan el delirio – en tanto creencia - con otras creencias que no se ajustarían a ciertos imperativos epistémicos claves. Por lo general, por ejemplo, las creencias racistas y clasistas, se sostiene, no son etiquetadas como delirantes, pero si tienden a ser inflexibles, indebidamente justificadas, resistentes a la evidencia y, por lo tanto, para muchos, irracionales. Desde un punto de vista de la observación externa del fenómeno, el delirio también poseería esas características, sin embargo, a las creencias racistas o clasistas no se les suele atribuir un carácter patológico. Para este grupo de autores el delirio podría ser localizado en un continuo con otros tipos de creencias que no parece instanciar ciertas propiedades epistémicamente deseables (Bortolotti & Miyazono 2015; Bortolotti 2023). Sin embargo, existe un debate respecto de donde y cómo definir la demarcación entre creencias anormales patológicas – como las delirantes - y no patológicas – como las creencias racistas y clasistas, entre muchas otras.

La discusión respecto de la demarcación del carácter patológico del delirio en la literatura se ha intentado relacionar con una discusión más amplia que ha sido predominantemente formulada en términos dicotómicos. Un marco dominante en la filosofía de la medicina tradicional ha caracterizado a aquello que denominamos ‘enfermedad’ o ‘patología’ como un concepto objetivo de carácter biomédico (Kendell 1975; Boorse 1977; Kingma 2013, 2017). Por otra parte, otra alternativa ha enfatizado la idea de que toda atribución de patología a un estado estaría basada en juicios de valor con importante influencia sociopolítica e incluso ideológica (Canguilhem 1970; Foucault 2005, entre otros). A la primera opción se le ha denominado naturalismo mientras que a la segunda se le ha denominado normativismo. En este capítulo exploramos algunos aspectos fundamentales de estos enfoques y especificamos algunas de sus debilidades. Luego de esto, indicamos que ninguna opción por sí sola logra ofrecer una respuesta satisfactoria a la pregunta sobre lo patológico del delirio y que, por esto, respuestas que combinen ambos aspectos del análisis de lo patológico podrían erigirse como una opción plausible. Finalmente, exploramos críticamente la aplicación de una propuesta híbrida clásica en la literatura contemporánea – la propuesta de la Disfunción Dañina (Wakefield 1992, 2007; Faucher & Forest 2021) – al caso del delirio (Miyazono 2015) y concluimos con algunos comentarios críticos sobre la forma en que el debate sobre lo patológico del delirio ha sido formulado en la literatura.

## 2. Naturalismo y Delirios

El naturalismo se fundamenta en la idea de que el normal funcionamiento de un organismo biológico – como los humanos - tiene como objetivo la supervivencia y la reproductividad; en consecuencia, un subsistema disfuncional dentro del organismo debiese entorpecer que ambos objetivos sean alcanzados (Boorse 1977). Acá, el concepto de subsistema se puede entender tanto a nivel físico, por ejemplo: órganos internos o sistema nervioso; como a nivel mental o cognitivo, por ejemplo: subsistemas de memoria, de lenguaje o de creencias. De este modo, una condición que se categoriza como patología, en general, sería una disfunción en un subsistema dentro de un organismo; a su vez, esta disfunción implicaría una desviación en el diseño biológico de la especie a la cual pertenece el organismo (Kingma 2017).

Para el naturalismo las patologías corresponden a una categoría natural, es decir, su identificación no depende de valores humanos (morales o políticos, por ejemplo), sino de criterios estadísticos y biológicos (Boorse 1977). Esta independencia valórica de lo que se comprende como patología sería la diferencia fundamental que tiene el naturalismo con el normativismo. Para que desde una mirada naturalista el delirio sea una creencia ‘patológica’ éste tendría que ser el resultado de una disfunción en el subsistema que origina creencias en un sistema humano (asumiendo un enfoque doxástico como lo hacen los enfoques dominantes en neuropsiquiatría clínica; ver López-Silva 2023). Esta disfunción, además, debiese ser comprendida como una condición libre de valores externos a la propia disfunción a la luz del diseño funcional de la especie.

Existen dos grandes críticas al naturalismo. La primera impugna la idea de que cualquier atribución de patológica debería estar libre de valores, es decir, ataca la idea defendida por naturalistas de que una disfunción que define a una patología no necesariamente es dañina o produce efectos negativos en los sujetos afectados. La segunda crítica surge en el contexto de la filosofía de la biología y pone en duda la existencia de las categorías naturales, esto es, la idea de que los conceptos de ‘salud’ y ‘enfermedad’ se pueden determinar objetivamente mediante información obtenida exclusivamente de las ciencias biológicas. Analicemos estas críticas a continuación.

Respecto a la primera crítica, históricamente se ha sostenido que cualquier patología debe producir algún mal en el sujeto afectado (Agich 1983; Cooper 2002). Privar una condición patológica de esta carac-



terística pareciera negar lo que hace relevante el discutirla. Esto, ya que si no fuera dañina o negativa para el sujeto ¿qué sentido tendría evitarla? Otras preguntas que pueden surgir en este contexto son: ¿por qué hay que intervenir en las enfermedades? o ¿por qué es preferible estar saludable antes que enfermo? Tales preguntas participan de un debate de índole moral y social, y, por lo tanto, normativo. Así, para algunos, tales preguntas pueden servir de evidencia de que en las discusiones en torno a condiciones patológicas la idea de lo ‘negativo para el sujeto’ ya cumpliría de antemano un rol fundamental (Kingma 2017). Se puede agregar a esta primera crítica lo complejo que puede resultar definir salud y enfermedad con criterios libres de valores. De acuerdo al naturalismo, la supervivencia de un organismo y su capacidad de reproducirse son objetivos que deben alcanzarse para que un organismo funcione bien y se considere ‘saludable’ (Boorse 1977; 2014). Pero el hecho de comprender estos objetivos como deseables, ya pareciera implicar una dependencia valórica al utilizarlos (Agich 1983; Kingma 2014). Más recientemente, se ha sugerido también las denominadas ‘categorías de referencia’ (Kingma 2007), esto es, grupos de la población identificados por edad, sexo y etnia que presentan todas los procesos y funciones con valores estadísticamente típicos también carecen de independencia valórica (Boorse 1977; 2014).

Kingma (2007) aclara el rol de la categoría de referencia con el siguiente ejemplo: “Tony Blair tiene buena salud si todas las partes y procesos que lo constituyen funcionan de tal manera que sean estadísticamente típicas en la categoría de referencia de hombre caucásico, de 53 años”. La autora argumenta que la selección de la edad, sexo y etnia para identificar las categorías de referencia no están justificadas empíricamente y, por esto, no se pueden desprender resultados objetivos e imparciales de la comparación por no estar libres de valores previos que la definen. De manera intuitiva tendría sentido la selección de edad, sexo y etnia para las categorías de referencia, pero no lo tendría si se agregara a ellas por ejemplo ser ciego o sordo. El hecho de evitar en las categorías de referencia condiciones como las señaladas, y sin justificar empíricamente por qué se hizo, es lo que hace sospechar que antes de definir siquiera lo que es salud y enfermedad ya estarían ambos conceptos interfiriendo. De este modo, definir salud y enfermedad a través de variables que supuestamente no están contaminadas de valores, cuando en realidad sí lo están de acuerdo a Kingma (2007) implica que el naturalismo cae en una circularidad.

En lo que concierne a la segunda crítica, no habría manera de definir lo que es salud y enfermedad con base en las ciencias biológicas (Cooper 2002; Ereshefsky 2009). Las características naturales o normales de una especie de acuerdo al naturalismo serían las que permitirían definir lo que se entiende por saludable. Pero al parecer esto no sería posible, y autores como Ereshefsky (2009) lo justifica desde tres áreas distintas de la biología. Desde la taxonomía biológica, se sostiene que esta no se dedica a definir características en común entre los miembros de una especie, sino más bien define el parentesco y conexión genealógica con otros miembros de la especie, la cual además está en permanente evolución por lo que difícilmente se podría hallar un diseño natural específico que un organismo de la especie debiese tener. Desde la genética, por otro lado, se afirma que es sabido que el genotipo compartido de dos individuos no se expresa fenotípicamente de igual manera cuando el entorno cambia. Esto dejaría en cuenta nuevamente lo complicado que sería poder identificar características de un diseño natural entre los miembros de una misma especie. Por último, desde la fisiología, Ereshefsky (2009) asegura que a pesar de que los manuales presentan descripciones detalladas de los organismos y sus mecanismos internos, lo cierto es que estos textos presentan descripciones idealizadas y sintetizadas de los órganos y sistemas de órganos. La finalidad de esto sería fijar un punto de partida para elaborar modelos más detallados de las partes de los organismos y sus mecanismos internos que puedan ajustarse mejor a la realidad (Wachbroit 1994).

Las críticas presentadas hasta aquí tienen en común la imposibilidad de evitar la intervención humana y sus valores en la conceptualización de la salud y la enfermedad. Si nos focalizamos solamente en la salud mental, las críticas parecieran cobrar incluso más sentido. Para el principal manual de diagnóstico en la actualidad - el DSM 5-TR -, la enfermedad o patología mental debería expresar una disfunción psicológica, biológica o del desarrollo (APA 2024). Sin embargo, los criterios diagnósticos para gran parte de los trastornos mentales contemplados en el manual no hacen referencia a procesos ni biológicos ni psicológicos que subyacen al trastorno (Aftab 2016). Por otra parte, el manual determina que, en ausencia de marcadores biológicos claros, no es posible identificar patologías basado solamente en criterios sintomáticos (al cuál pertenecería el delirio), por lo que se utiliza en su lugar el criterio general de malestar o discapacidad clínicamente significativos (APA 2024). Dicho de otro modo, hay trastornos mentales contemplados en el manual cuyos diagnósticos se realizan sin necesidad

de considerar un sustento biológico que los justifique, lo cual resulta ser problemático para la conceptualización naturalista de la patología y sobre todo, del delirio.

### 3. Normativismo y el Criterio de lo Patológico

En palabras generales, el normativismo – en su versión más básica - refiere a la idea de que la atribución de carácter patológico a una condición física o mental siempre implica un juicio de valor respecto del efecto negativo o perjudicial de la condición en la vida de la persona (ver, Faucher & Forest 2021). Esto, sin duda, dependerá de qué es exactamente aquello que los humanos consideramos negativo para el bienestar (Ereshfsky 2009; Amoretti & Lalumera 2021). En el normativismo, categorías como ‘saludable’ o ‘patológico’ no son valóricamente neutras; mientras la primera tiene que ver con estados físicos y mentales deseables, la segunda referiría a aquellos estados indeseables. El normativismo aplicado a la psiquiatría ha sido criticado sistemáticamente. El hecho de proponer que lo patológico es valórico-dependiente tendría consecuencias problemáticas.

Una primera crítica sostiene que lo que el normativismo considera patológico es dependiente en su totalidad de lo que la sociedad (o grupo de especialistas) piensan al respecto de una condición (Kingma 2017). De este modo, no se podría juzgar negativamente lo que en otra época fue considerado patológico, ya que lo que se considera negativo o dañino, al ser valórico es susceptible de ser modificado con el paso del tiempo. Así, los normativistas no podrían sostener, por ejemplo, que el hecho de calificar por tanto tiempo a la homosexualidad<sup>1</sup> como una patología sea un error que actualmente está subsanado; a lo más podrían afirmar que actualmente evaluamos a una condición de acuerdo a valores distintos a los del pasado. Una consecuencia de esta crítica es que nada impediría que en un tiempo futuro los valores con que se evalúan a una condición se inviertan; así, podría darse incluso el caso de que la homosexualidad vuelva a considerarse una condición patológica.

Una segunda crítica apunta al alcance de la propuesta. Pensemos en el caso de la salud en plantas y animales. En este caso, pareciese que una enfermedad podría ser atribuida significativamente sin necesidad

---

<sup>1</sup>Desde el DSM-III (1980) y sus ediciones posteriores no se incluye a la homosexualidad entre sus trastornos.

de que el ser humano realice una evaluación valórica negativa de la condición. Un jardinero, por ejemplo, podría considerar muy positivo que la maleza de su jardín se esté muriendo, lo cual no quita el hecho de que considere también que la misma esté enferma (Kingma 2017). De acuerdo a esta crítica, la patología de una condición al parecer puede darse de manera independiente de si existe o no una valoración negativa de la misma (tal como propone el naturalismo).

Una tercera crítica señala que el normativismo no siempre logra diferenciar con claridad la patología de otros infortunios. Muchas condiciones o situaciones que enfrenta una persona pueden ser dañinas para ella sin que se consideren patológicas. Respecto a la obesidad, por ejemplo, aún no hay consenso sobre si considerarla patología o no, a pesar de que todos concuerdan que no es una condición beneficiosa para la salud (Ereshfsky 2009). Por otra parte, a pesar de que la timidez y las inseguridades no sean consideradas patológicas, en general evitan que las personas afectadas puedan enfrentar situaciones que de no hacerlo saldrían perjudicadas.

Otro tipo de crítica - presentada principalmente por Bortolotti (2015; 2022) - se centra directamente en el caso de los delirios y sostiene que al menos en un grupo de estos, a pesar de que se pueden observar efectos dañinos, no necesariamente implica que sean causados por los delirios sino que por otras condiciones de base o por el cuadro clínico general. Dicho de otro modo, puede quedar en duda si los efectos dañinos atribuidos a un delirio sean realmente causados por este, y que tanto el delirio como los efectos dañinos que normalmente se asocian a él, son realmente una respuesta a otra condición que subyace a ambos. Los delirios motivados, llamados así por la presencia de influencias motivacionales en la formación de creencias (McKay et al. 2005; Bortolotti 2015), ayudarían a aliviar un malestar emocional, baja autoestima o los efectos adversos de una experiencia traumática. Este tipo de delirio suele describirse como mecanismos de defensa o incluso como funciones rasgo (McKay et al. 2005). En un caso presentado por Butler (2000), y también usado de ejemplo por otros autores, un sujeto a quien denomina B.X. sufre un accidente con severas lesiones cerebrales que lo dejan tetrapléjico. Un año después del accidente desarrolla la creencia delirante sobre la continuidad de la relación amorosa con su pareja anterior N, quien había cortado todo tipo de contacto con él luego del accidente. B.X estaba convencido incluso de que se había casado recientemente con N, a pesar de que ella en realidad ya estaba en una relación con otra persona. Este ejemplo, puede reflejar la manifestación de un mecanis-

mo de defensa psicológico motivado por la negación de la devastadora realidad en la cual su cuerpo se encuentra paralizado y N ya no lo ama. Tanto en este caso como en otros de similares características, pareciera que los delirios y las consecuencias dañinas que las acompañan son respuesta a otro problema de base, por lo que sostener que los delirios sean los causantes de los efectos negativos no pareciera tener mucho sentido. Es necesario mencionar que los casos donde se pueden aplicar estos mecanismos de defensa o de adaptabilidad se reduce sólo a un grupo de delirios, esto se debería a que no todos los delirios pueden explicarse satisfactoriamente mediante factores psicológicos (Lancelotta & Bortolotti 2019, Bortolotti 2022).

#### **4. Condiciones Patológicas como Disfunciones Dañinas: Un Problema Abierto**

Las secciones anteriores dejan ver que ninguna de las propuestas para describir qué es lo que hace patológico a una condición está libre de problemas. Lo anterior motivó la búsqueda de criterios híbridos que lograsen capturar la complejidad de la vida humana, y, sobre todo, de los conceptos de salud y enfermedad. Es así como el trabajo de Wakefield (1992a, 1992b) combina el normativismo con el naturalismo para indicar que aquello que etiquetamos como ‘patológico’ lo es en virtud de ser el producto de una disfunción dañina (harmful dysfunction). En este contexto, lo ‘daño’ captura el juicio de valor o consideración bajo el estándar sociocultural que forma parte de la categorización de un desorden como tal. Existirán ciertas normas, entonces, que dirigen la atribución de daño a una condición. Por otra parte, la dimensión ‘disfuncional’ de la propuesta refiere al aspecto objetivo que implica un fallo del funcionamiento normal del organismo (Wakefield 1992b; 2007). Mediante esta hibridación, Wakefield estaría en la posición de evitar varios de los problemas asociados al normativismo y el naturalismo cuando son aplicados por separado.

Una de las únicas aplicaciones de la propuesta de Wakefield al análisis de los delirios se encuentra en el trabajo de Miyazono (2015; 2019). El autor sostiene que los delirios – entendidos como creencias - son patológicos porque implican la existencia de cierto tipo de disfunciones que resultan ser perjudiciales para el sujeto. Lo que haría patológico al delirio, por lo tanto, sería su impacto negativo significativo en el bienestar del sujeto (elemento perjudicial) y el hecho de que algunos mecanismos o subsistemas directa o indirectamente relacionados con

su producción no cumplen las funciones originales para los que fueron seleccionados evolutivamente (elemento disfuncional). Respecto de este último punto, el foco dominante en la neuropsiquiatría actual se ha puesto en las alteraciones del sistema a cargo de la producción de creencias (Davies et al. 2001; Corlett et al. 2007; 2015; López-Silva 2023b).

Si bien la aplicación de la propuesta de la disfunción dañina al caso de los delirios parece plausible en primera instancia, no está exenta de importantes problemas. Uno de los problemas es señalado por Bortolotti (2022), quien señala que, muchas veces, el daño atribuido al delirio en realidad no es causado directamente por este, sino que por la condición de base que lo produce. Esta idea parece ser consistente con la propuesta que indica que el delirio sería una especie de respuesta adaptativa a ciertos funcionamientos problemáticos del apartado cognitivo humano (López-Silva 2023c; 2024). Sin embargo, parecen existir problemas mucho más profundos que éste. Por ejemplo, la propuesta de la disfunción dañina está pensada para analizar condiciones y no síntomas específicos. El error que comete Miyazono es exactamente éste, ignorar el contexto clínico en el cuál el delirio es reportado. Creemos que este error categorial se debe a, por lo menos, dos razones. Primero, el autor parece confundir el fenómeno delirante con sus características externas, algo ya señalado por Jaspers (1963). Segundo, el autor asume la continuidad epistémica entre delirio – entendidos como creencias – con otros tipos de creencias que no parecen instanciar las cualidades paradigmáticas de las creencias (por el ejemplo de las creencias racistas y clasistas). Este asunto parece impregnar toda la formulación del análisis del carácter patológico del delirio.

Asumir una continuidad epistémica entre delirio y creencias anormales no delirantes parece pasar por alto la idea de que el delirio se constituye como un quiebre en la experiencia de un mundo común (Fuchs 2015; Henriksen & Nilsson 2017; López-Silva 2024). La certeza con la que se reporta el delirio no surge de un proceso de razonamiento, aunque éste sea sesgado y conducido por la emocionalidad. La certeza del delirio es impuesta. Para el paciente delirante la realidad simplemente no puede ser de otra cosa por lo que asumir una continuidad de los criterios epistémicos para los delirios y las creencias anormales no delirantes parece ignorar todo esto. Este y otros problemas hacen que el estado actual de la aplicación de la propuesta de la disfunción dañina al análisis del carácter patológico del delirio se encuentre incompleta. Es de esperar que futuras aplicaciones tengan en cuenta no solo el contexto clínico en el cuál se reporta el delirio, sino que también la complejidad

de las transformaciones en la estructura de la conciencia que éste implica, y como esto parece desarticular la continuidad epistémica que sirve para fundamentar la formulación actual de problema.

## Referencias

Agich, G. J. (1983). Disease and value: A rejection of the value-neutrality thesis. *Theoretical Medicine*, 4, pp. 27–41.

Aftab, A. (2016). Mental Disorders and Naturalism. *American Journal of Psychiatry Residents' Journal*, 11, pp. 10-12. <https://doi.org/10.1176/appi.ajp-rj.2016.110304>.

Amoretti, C. & Lalumera, E. (2022). Wherein is the concept of disease normative? From weak normative to value-conscious naturalism. *Medicine, Health Care and Philosophy*, 25 pp. 47–60.

American Psychiatric Association [APA]. (2024). *Diagnostic and statistical manual of mental disorders*, 5th ed. (DSM-V-TR). Washington, DC: American Psychiatric Publishing. <https://doi.org/10.1176/appi.books.9780890425596>.

Bayne, T. & Pacherie, E. (2004). Bottom-up or Top-down. Campbell's rationalist account of monothematic delusions. *Philosophy, Psychiatry, & Psychology*, 11(1) , pp. 1-11.

Bayne, T. & Pacherie, E. (2005). In defence of the doxastic conception of delusion, *Mind & Language*, 20 (2), pp. 163–188.

Beer, M. D. (1996). Psychosis: A history of the concept. *Comprehensive Psychiatry*, 37(4), pp. 273–291. [https://doi.org/10.1016/S0010-440X\(96\)90007-3](https://doi.org/10.1016/S0010-440X(96)90007-3).

Berrios, G. E. (1991). Delusions as “wrong beliefs”: A conceptual history. *The British Journal of Psychiatry*, 159(14) , pp. 6–13.

Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford University Press.

Bortolotti, L. (2015). The epistemic innocence of motivated delusions. *Consciousness and Cognition*, 33 pp. 490–499. <https://doi.org/10.1016/j.concog.2014.10.005>.

Bortolotti, L. (2020). The epistemic innocence of irrational beliefs. Oxford: OUP. <https://doi.org/10.1093/oso/9780198863984.001.0001>.

Bortolotti, L. (2022). Are delusions pathological beliefs?. *Asian Journal of Philosophy*. <https://doi.org/10.1007/s44204-022-00033-3>.

Bortolotti, L. (2023). *Why delusions matter*. London: Bloomsbury. <https://doi.org/10.5040/9781350163331>.

Bortolotti, L. & Miyazono, K. (2015). Recent work on the nature and development of delusions. *Philosophy Compass*, 10(9) , pp. 636-645.

Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, 44 , pp. 542–573.

Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy*, 39(6) , pp. 683-724. <https://doi.org/10.1093/jmp/jhu035>.

Butler, P. (2000). Reverse Othello Syndrome Subsequent to Traumatic Brain Injury, *Psychiatry*, 63:1 , pp. 85-92. <https://doi.org/10.1080/00332747.2000.11024897>.

Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annual review of psychology*, 62, pp. 271–298. <https://doi.org/10.1146/annurev.psych.121208.131622>.

Corlett, P., et al. (2007). Disrupted Prediction-error Signal in Psychosis: Evidence for an Associative Account of Delusions. *Brain*, 130, pp. 2387–400.

Cooper, R. (2002). Disease. *Studies in History and Philosophy of Science Part c: Studies in History and Philosophy of Biological and Biomedical Sciences*, 33(2), pp. 263–282.

Davies M., Coltheart M., Langdon R., & Breen N. (2001). Monothe-matic delusions: towards a two-factor account. *Philosophy, Psychiatry, & Psychology*, 8(2), pp. 133-58.

Ereshefsky, M. (2009). Defining Health and Disease. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 40, pp. 221–227.

Faucher, L., & Forest, D. (Eds.). (2021). *Defining mental disorder: Jerome Wakefield and his critics*. The MIT Press. <https://doi.org/10.7551/mitpress/9949.001.0001>.



Fuchs, T. (2015). Pathologies of Intersubjectivity in autism and schizophrenia. *Journal of Consciousness Studies*, 22, p. 191.

Henriksen, M. G., & Nilsson, L. S. (2017). Intersubjectivity and psychopathology in the schizophrenia spectrum: complicated 'we', compensatory strategies, and self-disorders. *Psychopathology*, 50 pp. 321–333. <https://doi.org/10.1159/000479702>.

Jaspers, K. (1963). *General Psychopathology* (7th ed.). Manchester: Manchester University Press.

Kendell, R. (1975). The concept of disease and its implications for psychiatry. *British Journal of Psychiatry*, 127 pp. 305–15.

Kingma, E. (2013). Naturalist accounts of mental disorder, in *The Oxford Handbook of Philosophy and Psychiatry*. Edited by Fulford KWM, Davies M, Gipps R, Graham G, Sadler JZ, Stanghellini G, Thornton T. Oxford, United Kingdom, Oxford University Press, pp 363–384.

Kingma, E. (2007). What is it to be healthy? *Analysis*, 67 , pp. 128–133.

Kingma, E. (2017). Disease as Scientific and as Value-Laden Concept. In: Schramme, T., Edwards, S. (eds) *Handbook of the Philosophy of Medicine*. Springer, Dordrecht. [https://doi.org/10.1007/978-94-017-8688-1\\_75](https://doi.org/10.1007/978-94-017-8688-1_75).

Lancellotta, E., & Bortolotti, L. (2019). Are clinical delusions adaptive? *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(5), Article e1502. <https://doi.org/10.1002/wcs.1502>.

López-Silva, P. (2024). Thinking in Schizophrenia and the Social Phenomenology of Thought Insertion. *Philosophical Psychology*, <https://doi.org/10.1080/09515089.2024.2334720>.

López-Silva, P. & McClelland, T. (2023). *Intruders in the Mind: Interdisciplinary Perspectives on Thought Insertion*. Oxford: Oxford University Press.

López-Silva, P. (2023a). Creencias e Imaginaciones: Examinando el Problema Etiológico de los Delirios. *Pensamiento: Revista de Investigación e Información Filosófica*, 302, pp. 274-286.

López-Silva, P. (2023b). La etiología de los delirios psicóticos en la neuropsiquiatría actual. *Revista Ciencias de la Salud*, 21(2), pp. 1-16.

López-Silva, P. (2023c). Minimal Biological Adaptiveness and the Phenomenology of Delusions in Schizophrenia. In A. Falcato & J. Gonçalves (eds). *The Philosophy and Psychology of Delusions*. London: Routledge.

López-Silva, P., Núñez de Prado-Gordillo, M. & Fernández, V. (2023). What are delusions? Examining the typology problem. *WIREs Cognitive Sciences*, Online first, <https://doi.org/10.1002/wcs.1674>.

McKay, R., Langdon, R., & Coltheart, M. (2005). “Sleights of mind”: Delusions, defences, and self-deception. *Cognitive Neuropsychiatry*, 10(4), pp. 305–326. <https://doi.org/10.1080/13546800444000074>.

Miyazono, K. (2015). Delusions as harmful malfunctioning beliefs. *Consciousness and cognition*, 33, pp. 561-573. <https://doi.org/10.1016/j.concog.2014.10.008>.

Miyazono, K. (2019). *Delusions and beliefs: A philosophical inquiry*. Routledge.

Miyazono, K. (2022). Replies to critics. *Asian Journal of Philosophy*, 1(2). <https://doi.org/10.1007/s44204-022-00048-w>.

Wakefield, J.C. (1992a). The concept of mental disorder. On the boundary between biological facts and social values. *American Psychologist*, 47, pp. 373–388.

Wakefield, J.C. (1992b). The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis. *World Psychiatry*, 6(3), pp. 149-156. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2174594/>.

Wachbroit, R. (1994). Normality as a biological concept. *Philosophy of Science*, 61, pp. 579–591.

Woolfolk, R. (2001). The Concept of Mental Illness: An Analysis of Four Pivotal Issues. *The Journal of Mind and Behavior*, 22(2), pp. 161-178.

# Acerca del debate sobre el rol adaptativo de los delirios en la filosofía de la psiquiatría

Pablo López-Silva<sup>a</sup>, Channel Alarcón<sup>b</sup>, Tiare Quiroz<sup>b</sup>, Juan Riveros<sup>b</sup>,  
Martina Serrat<sup>b</sup>, Pablo Álvarez<sup>b</sup>, Álvaro Cavieres<sup>bc</sup>

## Resumen

Con una mayor prevalencia en la esquizofrenia, el delirio es considerado como la marca de la psicosis. Desde un punto de vista clínico, el delirio es predominantemente entendido como un tipo de creencia desadaptativa debido a sus consecuencias socio-funcionales para las diversas dimensiones de la vida de los pacientes. Sin embargo, durante los últimos años parece haber resurgido la idea de que el delirio podría tener un rol adaptativo. Este capítulo explora este debate. Primero, revisamos el marco de referencia conceptual que parece definir la formulación dominante del debate, esto es, la idea de que los delirios pueden ser comprendidos como un tipo de creencia. A esto se

---

<sup>a</sup>Escuela de Psicología, Universidad de Valparaíso, Chile. Hontaneda 2653, Of. 4.2., Valparaíso, Chile.

Contacto: pablo.lopez.silva@gmail.com

The Collaborating Centre for Values-based practice in Health and Social Care, St Catherine's College, Universidad de Oxford

Instituto de Sistemas Complejos de Valparaíso, Chile

Instituto Milenio para Investigación en Depresión y Personalidad, MIDAP, Chile

Centro de Investigación del Desarrollo en Cognición y Lenguaje, Universidad de Valparaíso, Chile

<sup>b</sup>Escuela de Psicología, Universidad de Valparaíso, Chile.

<sup>c</sup>PLS y AC agradecen el financiamiento del proyecto FONDECYT regular n° 1221058 otorgado por la Agencia Nacional de Investigación y Desarrollo (ANID). PLS agradece el auspicio del proyecto Renewing Phenomenological Psychopathology (Discretionary International Exchange Award / 223452/Z/21/Z) otorgado por la Wellcome Trust, UK a Matthew Broome y Giovanni Stanghellini.

le ha denominado el enfoque doxástico del delirio. Luego de esto, intentaremos clarificar las dos posiciones actuales del debate en torno al potencial valor adaptativo de los delirios. Finalmente, en la última sección terminaremos revisando algunos desafíos que parecen quedar abiertos cuando intentamos definir el rol adaptativo del delirio

**Palabras Clave:** Psicosis, Esquizofrenia, Delirios, Filosofía de la Psiquiatría

## Abstract

Delusions are widely regarded as a hallmark of psychosis, with a higher prevalence in schizophrenia. From a clinical perspective, delusions are predominantly understood as a form of maladaptive belief due to their socio-functional consequences on various aspects of patients' lives. However, in recent years, the idea that delusions may play an adaptive role appears to have resurfaced. This chapter explores this ongoing debate. First, we review the conceptual framework that seems to define the dominant formulation of the debate: the idea that delusions can be understood as a type of belief, a view commonly referred to as the doxastic approach to delusions. Following this, we aim to clarify the two current positions regarding the potential adaptive value of delusions. Finally, in the last section, we examine several challenges that remain unresolved when attempting to define the adaptive role of delusions.

**Keywords:** Psychosis, Schizophrenia, Delusions, Philosophy of Psychiatry

## 1. Introducción

El delirio es considerado como uno de los síntomas más complejos de la psicosis (American Psychiatric Association 2014; Bolding et al. 2012; Jaspers 1993). Con una mayor prevalencia en la esquizofrenia, el delirio parece llamar la atención de psiquiatras y filósofos por las maneras en que desafía nuestras comprensiones conceptuales y clínicas más fundamentales sobre la mente (López-Silva 2014; López-Silva y Cavieres 2022). Desde un punto de vista psiquiátrico-clínico, el delirio afecta diversas dimensiones del sujeto. Coltheart et al. (2011) señalan que los sujetos con esquizofrenia poseen dificultades para comunicarse de forma eficaz con otras personas en entornos sociales. Van Duppen (2017) analiza la existencia de una generalizada dificultad para comprender las situaciones sociales, lo que parece afectar la capacidad de sincronía e intencionalidad social de los pacientes. Kohler et al. (2003) menciona la existencia de problemas en la capacidad de reconocer las emociones del otro, sobre todo las emociones “negativas” como el miedo y la tristeza.

De esta forma, la incapacidad de establecer y/o mantener relaciones sociales y afectivas con otras personas (Motut et al. 2023) puede dificultar las interacciones sociales y familiares de los sujetos que experimentan delirio en esquizofrenia (Cavieres y López-Silva 2022; Fusar-Poli et al. 2022; National Institute of Mental Health 2015). Como consecuencia de las dificultades a nivel de interacción social, en algunos casos, los pacientes pueden llegar a formar actitudes de irritabilidad y hasta violencia (Báez 2012).

Enlistar los problemas y dificultades que presentan los pacientes delirantes en esquizofrenia en sus diversas dimensiones vitales tomaría mucho tiempo. El punto es que, a la luz de esta evidencia, la idea de que el delirio sería un estado ‘patológico’ y ‘desadaptativo’ parece ser intuitiva. Lo anterior exactamente por los malestares y fallas en funcionamiento cognitivo, emocional y social que parece involucrar. Esta posición dominante en la psiquiatría clínica parece entender lo ‘adaptativo’ como la propiedad de un estado o proceso de incrementar las probabilidades de supervivencia y reproducción de un organismo (R. T. McKay & Dennett 2009; Millikan 1993). Sin embargo, durante los últimos años, la posición dominante ha sido criticada bajo la premisa de que los delirios podrían ser adaptativos si se consideran una serie de factores motivacionales y contextuales (e.g., DeVlyder 2019; Garson 2021; Kingsbury 2006; Lancellotta y Bortolotti 2019 2020; López-Silva 2023).

Este capítulo explora el debate acerca del valor adaptativo de los delirios. Primero, revisamos el marco de referencia conceptual que parece definir la formulación dominante del debate, esto es, la idea de que los delirios pueden ser comprendidos como un tipo de creencia. A esto se le ha denominado el enfoque doxástico del delirio. Luego de esto, intentaremos clarificar las dos posiciones actuales del debate. Finalmente, en la última sección, terminaremos revisando algunos desafíos que parecen quedar abiertos cuando intentamos definir el rol adaptativo del delirio.

## 2. Delirios y Doxasticismo

### El problema Tipológico

La principal puerta de entrada al fenómeno delirante es aquello que los mismos pacientes reportan. Con sus intrínsecas dificultades metodológicas y epistemológicas, los reportes verbales de los pacientes parecen ser el medio principal para obtener información inicial sobre el delirio (López-Silva y Cavieres 2023). Ahora bien, cuando Geraldine

comenta que su gin & tonic quedó ‘muy amargo’ ¿qué está reportando Geraldine? No es problemático pensar que este reporte verbal adquiere sentido como el reporte de una experiencia sensorial gustativa. De igual forma, cuando Geraldine comenta que le hubiese gustado haber sido una técnico en radares de aviones durante la II Guerra Mundial igual que su Abuela, el reporte verbal de Geraldine es el reporte de un estado cognitivo imaginativo (con algunos toques de deseo también). Cuando Geraldine me dice que su nuevo auto es gris, podemos entender ese reporte como el reporte de una experiencia sensorial de tipo visual.

¿Pero qué ocurre cuando un paciente reporta que los árboles les han estado insertando pensamientos en su cabeza? (Payne 2013) ¿es acaso esta expresión el reporte de una experiencia sensorial? ¿Es una explicación metafórica de un estado simplemente confuso? ¿Es algo que está imaginando? Como he mencionado en otros lugares, esto es un problema en la psicopatología actual ya que implica el desafío de identificar el tipo de estado mental que está reportando un paciente delirante (López-Silva 2023a). Así, a este asunto se le ha denominado el problema tipológico de los delirios. El principal objetivo de las alternativas a este problema será “definir el tipo de estado mental que se ve instanciado en los reportes de pacientes de los cuáles decimos que sufren de delirios” (López-Silva y Cavieres 2022, p. 74).

Esta formulación del problema tipológico podría ser criticada por ser excesivamente conceptual, un asunto más importante para los filósofos que para los psiquiatras. No estamos completamente de acuerdo con esto. Tal como señalan López-Silva y Cavieres (2022, p. 75), “toda la investigación experimental y el tratamiento psicoterapéutico de los delirios descansa en la idea de que este fenómeno es un tipo específico de estado mental” (ver también Bayne y Pacherie 2005). De esta manera, si los delirios son entendidos como una instanciación de un tipo de estado mental M, la investigación experimental sobre los delirios podría focalizarse en cómo el tipo de estado mental M se produce en el cerebro y cómo las alteraciones en los procesos de producción de M producirían estados delirantes en un sujeto (Coltheart et al. 2011; López-Silva 2022). Sin la definición tipológica M sería muy difícil avanzar en la investigación empírica y clínica de los delirios; ésta última es sólo posible si entendemos cómo lidiar con un tipo de estado mental específico cuando los pacientes se presentan en la clínica buscando ayuda. Si no comprendemos qué tipo de estado mental es un delirio, y por ende no entendemos sus características principales ni su comportamien-

to paradigmático, ¿cómo podríamos generar formas de afrontamiento psicoterapéutico que solamente son eficaces y eficientes si conoces el tipo de estado mental con el que estamos lidiando y su comportamiento paradigmático? Así, tal como señalan López-Silva y Cavieres (2022):

Gran parte de la psicoterapia actual implica anteponerse y prever potenciales comportamientos del delirio en un paciente; esto sería muy difícil de lograr sin tener una respuesta al problema tipológico, y por esto este debate no constituye un mero ejercicio teórico, sino que un problema práctico para la psicopatología actual también (75).

### **Potenciales alternativas al problema Tipológico**

El enfoque dominante en el debate tipológico que fundamenta la principal formulación del problema acerca del valor adaptativo del delirio es el denominado enfoque doxástico. El doxasticismo sobre los delirios propone que los reportes delirantes instancian creencias anormales (Bayne y Pacherie 2005; Bortolotti 2010, 2012; López-Silva 2017). Para este enfoque, los delirios son creencias que carecen de ciertas características paradigmáticas, pero que con todo, pueden considerarse como creencias (Bayne 2010). Una de las fortalezas de este enfoque radica en el hecho de que los delirios parecen ser reportados de la misma forma en que son reportadas algunas creencias, esto es, con variables pero altos grados de certeza subjetiva (Green et al. 2018). Existen al menos dos grandes formulaciones dentro del modelo doxástico de los delirios. Por una parte, el enfoque *top-down* sugiere que ciertos patrones alimentados durante el tiempo por ideas delirantes comienzan a sedimentarse en la forma en que un sujeto comienza a interpretar la realidad. Luego este patrón delirante podría contaminar el contenido de la experiencia sensible de un sujeto, lo que implicaría la génesis del delirio (Campbell 2001). Por otra parte, la formulación *bottom-up* sugiere que la base de la creación de delirios *qua* creencias es la existencia de anomalías experienciales con contenido altamente anormal (Bayne y Pacherie 2005). Si bien existen otras distinciones dentro de la formulación *bottom-up* del modelo, el enfoque doxástico se ha convertido en la alternativa dominante dentro de la discusión tipológica por su claridad conceptual, alcance empírico y poder heurístico (Bortolotti 2010, 2020).

Si bien el enfoque doxástico es considerado el dominante en el contexto de la filosofía analítica de la mente y las ciencias cognitivas, existe

otro grupo de teorías que nacen como crítica a este. El antidoxasticismo referirá a un grupo de teorías que indican que los delirios no parecen cumplir con los requerimientos para ser considerados creencias (ver por ejemplo Schwitzgebel 2012). Un ejemplo de este tipo de acercamiento al problema tipológico lo encontramos en la propuesta Imaginística de los delirios, la que agrupa a una serie de autores que asocian los delirios con alteraciones de la imaginación. Currie (2000) indica que los delirios son ‘alucinaciones cognitivas’, esto es, estados imaginativos mal identificados como creencias de un sujeto. Por su parte, McGinn (2004) sugiere que los delirios son estados imaginarios creados por el descontrol y falta de monitoreo de la actividad imaginativa de un sujeto. En una línea similar, Egan (2008) indica que dado que los delirios comparten características de las imaginaciones y las creencias, los delirios deberían ser catalogados como *bimagnitions*. Sin embargo, esta idea ha recibido diversas críticas por su falta de claridad conceptual y falta de atractivo fenomenológico y empírico (Bortolotti 2010; López-Silva 2017). En cualquier caso, en las siguientes secciones intentaremos comprender el debate acerca del valor adaptativo del delirio, el cuál predominantemente se ha formulado en términos doxásticos.

### 3. El Delirio como Estado Desadaptativo

Además de estar fundamentada en el doxasticismo, la posición dominante en el debate acerca del rol adaptativo del delirio que revisaremos en este apartado está fundamentada en la idea de que los delirios son el subproducto de diversos déficits en el funcionamiento cognitivo y sensorial del sujeto (Coltheart et al. 2011; Davies y Coltheart 2023; López-Silva y Cavieres 2023; Sterzer et al. 2018). Por esto, la discusión acerca de lo ‘adaptativo’ y lo ‘patológico’ parece relacionarse continuamente en la literatura actual. Nosotros nos enfocaremos en lo ‘adaptativo’ desde un punto de vista doxástico.

En el clásico artículo ‘The Evolution of Misbeliefs’ McKay y Dennett (2009) sugieren que el ser humano ha evolucionado para preservar un sistema enfocado en formar creencias verdaderas. Esto le permitiría actuar eficazmente en el mundo incrementando sus chances de supervivencia. De manera fundamental, los autores indican que las creencias adaptativas *qua* verídicas deben ser fundamentadas, es decir, deben estar apropiadamente explicadas desde la evidencia y creencias pre-existentes. Desde este punto de vista, las creencias verdaderas serían



adaptativas porque aumentan la probabilidad de supervivencia por permitir comportamientos adaptativos eficaces en el punto.

Pues bien, basados en estas ideas, Coltheart et al. (2011) desarrollan su modelo de 2 factores de formación de creencias para explicar la formación de delirios *qua* creencias. En esta teoría, el primer factor es descrito como una deficiencia neuropsicológica en el procesamiento perceptivo, la cual determina el contenido de la creencia delirante. Mientras que el segundo factor corresponde a una alteración en la evaluación de las creencias, siendo independiente de su contenido; explicando el mantenimiento del delirio. De esto, Davies y Coltheart (2023) concluyen que si este sistema funciona de manera incorrecta, formando creencias erróneas, estas deben ser consideradas desadaptativas. Lo anterior debido a que el contenido falso de un delirio provocaría la incapacidad del sujeto para adaptarse adecuadamente a su entorno a la luz del desajuste epistémico entre creencia y realidad. La idea esencial acá es que una creencia incrementa las posibilidades de supervivencia en virtud de su veracidad.

En la misma línea, los modelos de Coltheart et al. (2011), McKay (2012), Davies y Coltheart (2023) plantean que en esquizofrenia, el funcionamiento del individuo es interrumpido por el grado de desajuste entre el contenido de la creencia delirante y la realidad experimentada por aquellos que no están delirando. En otras palabras, mientras exista una barrera entre la realidad experimentada por quienes no deliran y la creencia delirante de la persona, el delirio no cumpliría las condiciones necesarias para ser adaptativo.

Cabe señalar que desde una perspectiva lingüística, Berrios (1991) afirma que el contenido del delirio carece de una base probatoria, correspondiendo a una afirmación sin sentido (*empty speech act*). Por esto, el tipo de análisis propuesto por el debate que acá tratamos sería poco útil. Para Berrios (1991), si la persona delirante refiere que “es el enviado de Dios”, el contenido de dicha premisa no podría ser probado mediante la evidencia; esto porque sería un acto carente de contenido probatorio en sí mismo. En este sentido, Berrios (1991) menciona que el delirio corresponde a un acto de habla vacío disfrazado de creencia. Parece desprenderse de este enfoque la idea de que el delirio no sería un estado adaptativo de ninguna forma; sin embargo, esto no es del todo claro puesto que el planteamiento de Berrios (1991) se aleja de las condiciones epistémicas que impone el enfoque doxástico en el debate.

## 4. Delirio como Acto Adaptativo

Garson (2021) menciona que comúnmente se ha comprendido que un rasgo disfuncional implica la no realización de sus funciones. Por ejemplo, que los ojos no puedan ver producto de la ceguera, o bien que sólo puedan realizar su función de manera deficiente. Sin embargo, a la vez, Garson (2021) señala que “el mero hecho de que un rasgo no pueda realizar su función no lo convierte en disfuncional” (339); la incapacidad de un rasgo para realizar su función a un ritmo aceptable es necesaria pero no suficiente para clasificarlo como disfuncional (Garson 2021). Consideremos el caso de que los ojos están vendados. Garson en este caso reconoce que con la venda sus ojos no pueden cumplir la función de ver; no obstante, no son disfuncionales. Si aplicamos esta idea al caso del delirio, uno podría decir que el fenómeno cumple una función en esquizofrenia; esto es, explicar experiencias altamente anormales lo cual proveería una sensación de resolución y claridad (Fusar-Poli et al. 2022). Los delirios en esquizofrenia surgen como la última etapa de un período que involucra la gradual rarificación de la realidad, los otros y el sí-mismo (Conrad 1997; Feyaerts y Sass 2024; Fuchs 2005; Sass y Feyaerts 2024). En este período de ‘atmósfera delirante’, el delirio surgiría como una forma de reordenar la estructura de la conciencia (López-Silva 2023b; López-Silva y Abarca 2023). Por lo tanto, podríamos identificar una función siendo desplegada efectivamente. Ahora, si bien lo esperable es que la mente (sistema doxástico al menos) formule otro tipo de respuestas frente a los fenómenos extraños experimentados durante el pródromo de la esquizofrenia, para Garson (2021) no es posible identificar argumentos suficientes para clasificar al fenómeno del delirio en sí mismo como un rasgo disfuncional que es finalmente lo que lo haría desadaptativo.

Argumentando de manera diferente, Millikan (1986) sugiere que la condición de veracidad de una creencia carece de valor para determinar el ajuste adaptativo de un organismo a un entorno. El criterio principal para definir esto sería la funcionalidad. Millikan (1986) indica que una de las funciones propias de la creencia es “participar en inferencias para producir otras creencias” y que “ayuden a producir el cumplimiento de los deseos” (8-9). Al respecto, el autor sugiere que en condiciones especiales no importaría la veracidad de la creencia producida en un contexto anormal, esto porque igual podría estar cumpliendo los criterios recién mencionados. Acá el miedo y la preocupación que emergen por las experiencias sensoriales anómalas de la persona en contexto pre-

delirantes podrían definir el contexto de adecuación de la formación de un delirio en tanto función. Entonces, desde la perspectiva de Millikan (1986), por más que fueran falsas dichas creencias, si cumplen con la adecuación del sujeto en un contexto específico serían adaptativas. Acá la veridicalidad de la creencia formada jugaría un rol secundario. Por ejemplo, producir una creencia frente a tales experiencias sensoriales permite el surgimiento de nuevas inferencias, como creer que dada la confabulación de la sociedad en su contra (en un caso de delirio paranoide), en su casa pueda estar más seguro y protegido. Sin duda, esta posición invita a repensar el rol que las condiciones epistémicas de una creencia juegan en el debate acerca de los delirios, ¿sin embargo?, uno podría indicar que incluso si se motivan nuevas inferencias, tales inferencias podrían llegar a ser problemáticas (como fundamento para las alteraciones indicadas en la introducción).

Siguiendo con la exploración del concepto de adaptabilidad en el caso de los delirios, Lancellotta y Bortolotti (2019) indican que algunos delirios serían adaptativos desde un punto de vista psicológico, mientras que otros podrían serlo biológicamente. Según las autoras, un rasgo adaptativo debe ser comprendido en el contexto del entorno en el que se desarrolla. Tales rasgos podrían ser adaptativos en unos ambientes y desadaptativos en otros (Lancellotta 2022a, 2022b; Lancellotta y Bortolotti 2019, 2020). Parte de la propuesta de Lancellotta y Bortolotti (2019) se nutre de la teoría del pasador doxástico de seguridad propuesta por McKay y Dennett (2009). La idea es que existiría un sistema diseñado para activarse con el fin de proteger sus elementos de mayor valor. Cuando algún elemento de este sistema falla, el sistema deja de funcionar normalmente. Específicamente, los pasadores de seguridad doxásticos han sido producidos evolutivamente para abrirse como ‘llaves de paso’ cuando la persona se enfrenta a un estrés psicológico extremo, permitiendo la formación y mantenimiento de creencias anormales tales como los delirios (McKay y Dennett 2009). Lancellotta y Bortolotti (2020) concluyen que el mantenimiento de las creencias erróneas como fruto del colapso del pasador de seguridad protege el funcionamiento cognitivo de una persona aliviando su ansiedad frente a las experiencias anómalas.

Lancellotta y Bortolotti (2020) también describen los criterios básicos para considerar a una creencia como biológicamente adaptativa, los cuales tienen que ver con el mejoramiento de las posibilidades de supervivencia y reproducción de un individuo (ver introducción). Un enfoque que se ha hecho parte del debate respecto a la adaptabilidad biológica

corresponde al del procesamiento predictivo (Fineberg y Corlett 2016; Sterzer et al. 2016, 2018). Este enfoque concibe la arquitectura funcional del cerebro de manera jerárquica que tiene el objetivo de maximizar la evidencia de su modelo del mundo a través de una lógica Bayesiana. La idea es que el cerebro compara creencias previas (*Priors*) con los datos sensoriales disponibles (*Likelihood*) para computar respuestas. A la discrepancia entre un evento y un estímulo inmediato se le denomina error de predicción (Sterzer et al. 2018). La inferencia del modelo se puede maximizar a través de la inferencia activa al actuar sobre el mundo seleccionando evidencia sensorial para minimizar los errores de predicción (Sterzer et al. 2018).

Los autores Hemsley y Garety (1986) presentan el primer análisis explícitamente bayesiano del delirio, sugiriendo cómo la creencia, la evidencia y su interacción interrumpida podrían ser generadoras de inferencias aberrantes. Los sujetos dirigirían su atención y aprenderían sobre eventos comúnmente ignorados por otros, constituyendo la base para alucinaciones y delirios (Hemsley y Garety 1986). Sterzer et al. (2018) argumentan que los errores de predicción aberrantes actúan como impulsores en la formación de delirios. Fundamentalmente, las creencias delirantes se formarían como la mejor explicación a un error de predicción ruidoso e incierto (Sterzer et al. 2018). Acá, más que un problema o una disfunción en sí mismos, los delirios serían una respuesta imperfecta a una situación comprometida desde un punto de vista biológico y/o psicológico. En este contexto, los delirios serían biológicamente adaptativos por los beneficios epistémicos y funcionales.

Como fue señalado anteriormente, Lancellotta y Bortolotti (2019, 2020) mencionan la existencia de tipos de adaptabilidad. Respecto a la adaptabilidad psicológica de los delirios, mencionan que una creencia es psicológicamente adaptativa siempre y cuando mejore el bienestar, el propósito en la vida y el buen funcionamiento de la persona. Esto principalmente debido a que frente a la angustiante y desesperante experiencia del pródromo, la aparición del delirio aliviaría la incertidumbre. Un ejemplo presentado por las autoras corresponde a la anosognosia, donde individuos que experimentan la parálisis de una extremidad niegan su deterioro afirmando que sí pueden moverla (Lancellotta y Bortolotti 2019). A raíz de esto, Lancellotta y Bortolotti (2019) mencionan que cuando las personas necesitan vivir con las consecuencias del trauma, negar su discapacidad podría generar una adaptabilidad a corto plazo. Lancellotta y Bortolotti (2019) utilizan parámetros relevantes para referirse a la adaptabilidad psicológica de los delirios, mencionando que

los beneficios de las creencias erróneas varían según su contenido y el contexto en que se adquieren. Adicionalmente, utilizan la variable temporal para concluir que los delirios son psicológicamente dañinos y biológicamente desadaptativos a largo plazo, pero que ofrecerían beneficios a corto plazo, siendo en algún punto adaptativo. Lo anterior podría corresponder a una variable respecto a las características que podrían constituir a un rasgo como adaptativo. Hasta acá, es claro que en el debate, el concepto de lo 'adaptativo' es utilizado de manera variada y ambigua. Esto parece problemático y justifica una exploración del concepto como tarea preliminar para luego ponderar el atractivo argumentativo de las posiciones en el debate.

## 5. ¿Son Adaptativos los Delirios?

La presente contribución no ha pretendido resolver la pregunta que titula la actual sección. En estricto rigor, nos hemos dedicado a una labor preliminar, esto es, revisar en qué consiste el debate acerca del rol adaptativo del delirio y examinar las bases de su actual formulación en la literatura en filosofía de la mente y ciencias cognitivas. Hemos podido distinguir las 2 posiciones que parecen sobresalir en el estado actual del debate. Acá es inevitable observar una serie de desafíos y asuntos conceptuales que parecen quedar abiertos en la discusión actual. Creemos que la discusión de estos asuntos podría aportar al avance en el debate. Distingamos pues entre 2 tipos de problemas. El primer tipo tiene que ver con asuntos que parecen problemáticos (o al menos poco claros) dentro de cada posición en el debate. El segundo grupo tiene que ver con las ideas a la base de la formulación misma del problema acerca del valor adaptativo del delirio.

Revisemos el primer grupo de problemas. Primero, la posición desadaptativa de autores como McKay y Dennett (2009) no parece considerar la idea de que podrían existir creencias no verídicas que podrían cumplir un rol adaptativo en la vida mental y social de un sujeto. Imagine-mos que H posee una autoimagen sumamente negativa de él. Tanto su comunidad como la evidencia parece demostrar que H es una persona altamente eficiente y hasta es un ejemplo para otros. Acá, el grupo de creencias autorreferentes negativas en torno a la identidad de H parecen ser falsas. Sin embargo, es la negativa autoimagen de H lo que lo motiva a ser altamente productivo en su trabajo como una forma de compensar por esa autoimagen. En este caso, es en virtud de su falsedad que este grupo de creencias cumple una función que podría ser considerada

altamente adaptativa. Si bien es cierto que altos niveles de autoengaño negativo pueden ser dañinos, niveles saludables de autoengaño podrían precipitar comportamientos socialmente funcionales y, por lo tanto, incrementar las posibilidades de sobrevivencia de un sujeto. Obviamente, acá el problema será determinar los límites de aquellos niveles. Sin embargo, nuestro punto es que es posible pensar en creencias falsas que cumplan un rol adaptativo; por lo tanto, la veracidad de una creencia sería solamente un aspecto ortogonal para entender la naturaleza adaptativa o desadaptativa del delirio, no una condición suficiente.

Por otro lado, posiciones como la de Millikan (1986; 1993) que efectivamente consideran la insuficiencia del argumento de la veracidad de una creencia para determinar su nivel adaptativo parecen olvidar que después de todo una creencia verdadera podría ser más adaptativa que una falsa en muchos contextos. Imagina que H tiene la creencia (en el lugar T hay comida y agua fresca). Ciertamente, esta creencia podría ser adaptativa en virtud de que sea verdadera. En este sentido, la creencia funciona efectivamente como un estado adaptativo porque es verdadera. Por lo tanto, desestimar completamente el rol de la veracidad al determinar el valor adaptativo de las creencias parece encontrar otro tipo de contraejemplos que debilitaría la posición. Si bien el sistema doxástico podría sacrificar niveles de ajuste epistémico con la realidad bajo ciertas circunstancias (como por ejemplo en los periodos de atmósfera delirante), la mantención de un equilibrio —incluso homeostático— sería fundamental para transitar la realidad con niveles funcionales de creencias falsas funcionales y creencias verdaderas que incrementen la adaptación de un sujeto a su medio ambiente.

Por otra parte, la trinchera que defiende la posición adaptativa de los delirios también parece tener ciertas dificultades. Uno podría aceptar la idea de que el delirio surge como una manera de explicar estados experienciales altamente anormales y perturbadores. Los árboles insertando una serie de reglas en mi mente podría al fin explicar toda la rareza con la cual el mundo se me presentó por un largo tiempo (ver López-Silva 2018; Payne 2013). Esto parece ser incluso consistente con las descripciones fenomenológicas de la esquizofrenia (Fusar-Poli et al. 2022). Sin embargo, es inevitable preguntarse por qué los pacientes explican exactamente de esta forma la realidad. Si los delirios son un tipo de creencia, no es claro el origen de tales hipótesis explicativas para tales experiencias anómalas (Parrott 2021). Acá, el asunto no es que el acto de explicar sea problemático. Ciertamente, explicar la experiencia inmediata es uno de los aspectos más primitivos de nuestra

relación con el mundo y el sí-mismo (Guidano 1987, 1991). El problema es efectivamente el origen de hipótesis explicativas tan bizarras y porque tales hipótesis serían finalmente elegidas como las mejores opciones para explicar la experiencia. Acá, algunos podrían indicar que en realidad, ante la ausencia de otras opciones, el sujeto inevitablemente explica la realidad de esa manera, lo que llevaría a la sensación de imposición con la cual el delirio es experimentado en esquizofrenia. El problema con esta sugerencia es que simplemente no se puede falsear. ¿Cuáles son los medios por los cuáles podríamos determinar que el sistema doxástico delirante no posee otras opciones cuando intenta explicar experiencias anómalas? Por lo demás, esa sugerencia tampoco parece explicar el contenido mismo del delirio, el cual comúnmente toma formas extravagantes. ¿Por qué la única opción explicativa poseería ese nivel de extravagancia? Estos son asuntos centrales en este grupo de teorías que deberían ser profundizados.

Pasemos ahora al grupo de problemas que se desprende de la formulación misma del debate acerca de la naturaleza adaptativa del delirio. El primer problema tiene que ver con el hecho de que no es del todo clara la relación entre adaptabilidad psicológica y adaptabilidad biológica. Uno podría indicar que existen estados psicológicamente positivos —incluso que son placenteros— que no son adaptativos. Por ejemplo, Nesse (1998) indica que bajos niveles de ansiedad —que son psicológicamente placenteros— serían el producto de malfunciones etiológicas. La ansiedad tiene un rol adaptativo importante y este tipo de fallas, si bien generan estados psicológicos placenteros, implicaría una falla en el cumplimiento del rol mismo de la ansiedad. Uno podría también sugerir que existen estados que cumplen su función adaptativa exactamente por su carácter egodistónico. Dos ejemplos de estos son el dolor y la fiebre. En ambos casos, podemos ver ejemplos de estados que son psicológicamente indeseables, pero que cumplen su función exactamente en virtud de esa propiedad. Aparentemente, la distinción entre lo psicológico y lo biológico parece implicar que ambas podrían ser analizadas de manera separada. Pero, ¿es esto realmente posible? Las actuales corrientes postcognitivas (e.g., Bruin et al. 2018; De Haan 2020; Di Paolo et al. 2017; Maiese 2021; Nielsen 2023) han enfatizado la naturaleza corporeizada, contextualizada, enactiva y ecológica de la mente humana. Los denominados ‘estados psicológicos’ entonces podrían ser entendidos como una función de las dinámicas biológicas de un organismo. Estados psicológicos y estados biológicos serían parte de la continuidad de la vida en términos de complejidad funcional en los humanos. Sin duda,

la distinción psicológica/biológico podría informar niveles específicos de argumentación; sin embargo, para jugar un rol más determinante, mayor claridad se necesita al establecer la relación entre ambos niveles. ¿Por qué lo psicológicamente adaptativo no podría ser considerado biológicamente adaptativo y viceversa? Este es un asunto que merece ser tratado exhaustivamente en la literatura.

Acá llegamos a uno de los problemas más fundamentales en la formulación del debate acerca del rol adaptativo de los delirios. Toda la discusión parece formularse en torno al concepto de ‘rasgo adaptativo’. La idea es considerar si el delirio pudiese ser o no considerado como este tipo de rasgo, esto es, un rasgo adaptativo de la mente humana. El problema acá es que la noción misma de ‘rasgo adaptativo’ parece ser objeto de discusión en el campo de la filosofía de la biología evolutiva. La formulación actual del debate parece ser influenciada por el Darwinismo Clásico, que propone fundamentalmente que todos los organismos vivos comparten un ancestro en común y que la selección natural es la respuesta al cambio evolutivo, así como el único proceso que puede provocar adaptación (Pigliucci & Müller 2010). Un rasgo adaptativo sería elegido por la selección natural. Sin embargo, Raup (1992) afirma que el mecanismo propuesto por Darwin (1987) en beneficio de la competencia de las especies es insuficiente, siendo más bien un enfoque sobre las causas de las extinciones. El autor también plantea que la desaparición de cierta especie no responde únicamente al mecanismo de la selección natural, dado que en las extinciones masivas “las especies se extinguen no debido a su relativo éxito adaptativo en el escenario ecológico, sino ante todo porque demuestran estar mal equipadas para hacer frente a cambios abruptos en su entorno físico” (Gayon 2009, p. 336). Por su parte, Gould (2002 citado en Gayon 2009) rechaza la idea de que la selección natural sea el único medio de regulación de los procesos evolutivos, afirmando que esta no solo opera a nivel de organismos individuales (agencia), sino que comprende a grupos y especies. Gould indicará que dicho proceso no representa la totalidad de las causas de los cambios evolutivos (eficacia) y que el proceso de macroevolución es más que cambios sutiles intraespecíficos (alcance).

Si bien el párrafo anterior no logra hacer justicia a la complejidad de la discusión, lo que intenta mostrar es que incluso los conceptos más fundamentales utilizados en las formulaciones actuales del debate acerca del rol adaptativo del delirio parecen acarrear importantes problemas en los términos utilizados. Además de esto, el darwinismo clásico no parece ser dominante en la actualidad, donde teorías como el



neodarwinismo, la síntesis evolutiva extendida o la teoría neutralidad de la evolución parecen comprender de distintas maneras el concepto mismo de rasgo adaptativo y las formas en que tales rasgos son seleccionados evolutivamente para operar en el proceso de adaptación. Sin duda, explorar este tipo de diferencias terminológicas podrían informar de manera importante el debate acerca de si los delirios podrían ser considerados un rasgo adaptativo o no.

## 6. Conclusión

Este capítulo ha intentado explorar las bases conceptuales de la actual formulación del debate acerca de si los delirios tuvieran algún rol adaptativo. Si bien la visión desadaptativa parece ser dominante en la psiquiatría clínica, durante los últimos 5 años, una serie de autores la han desafiado apelando a la forma en que la experiencia delirante incluiría sensación de completitud, reducción de ansiedad y epifanías que lograrían reestructurar el campo fenoménico de la conciencia de un sujeto. Sin embargo, no solo en las bases de la formulación del problema, sino que también en las posiciones identificadas, existe una serie de conceptos problemáticos. Es de esperar que la comunidad pueda progresar en el debate, manteniendo el rol determinante de la evidencia médica, pero también integrando los reportes de aquellos que experimentan estos fenómenos desde la primera persona. Ciertamente, todo análisis conceptual debería integrar armoniosamente estos aspectos del fenómeno psicopatológico.

## Referencias

- American Psychiatric Association (Ed.). (2014). *Guía de consulta de los criterios diagnósticos del DSM-5*. American Psychiatric Publishing.
- Báez, J. (2012). El delirio y el discurso: Débiles referentes para arbitrar en la salud mental. *Tesis Psicológica*, 7, pp. 18-39.
- Bayne, T. (2010). Delusions as Doxastic States: Contexts, Compartments, and Commitments. *Philosophy Psychiatry & Psychology*, 17(4) pp. 329-336. <https://doi.org/10.1353/ppp.2010.0030>
- Bayne, T., & Pacherie, E. (2005). In Defence of the Doxastic Conception of Delusions. *Mind and Language*, 20(2), pp. 163-188. <https://doi.org/10.1111/j.0268-1064.2005.00281.x>

- Berrios, G. E. (1991). Delusions as “Wrong Beliefs”: A Conceptual History. *British Journal of Psychiatry*, 159(S14) , pp. 6-13. <https://doi.org/10.1192/S0007125000296414>
- Bolding, M. S., Lahti, A. C., Gawne, T. J., Hopkins, K. B., Gurler, D., & Gamlin, P. D. (2012). Ocular Convergence Deficits in Schizophrenia. *Frontiers in Psychiatry*, 3, p. 86. <https://doi.org/10.3389/fpsy.2012.00086>
- Bortolotti, L. (2010). *Delusions and other irrational beliefs*. Oxford University Press.
- Bortolotti, L. (2012). In Defence of Modest Doxasticism About Delusions. *Neuroethics*, 5(1) pp. 39-53. <https://doi.org/10.1007/s12152-011-9122-8>
- Bortolotti, L. (2020). *The Epistemic Innocence of Irrational Beliefs* (1.a ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780198863984.001.0001>
- Bruin, L. de, Newen, A., & Gallagher, S. (Eds.). (2018). *The Oxford Handbook of 4E Cognition*. Oxford University Press.
- Campbell, J. (2001). Rationality, meaning, and the analysis of delusion. *Philosophy Psychiatry & Psychology*, 8(2-3) , pp. 89-100. <https://doi.org/10.1353/ppp.2001.0004>
- Cavieres, Á., & López-Silva, P. (2022). Social Perception Deficit as a Factor of Vulnerability to Psychosis: A Brief Proposal for a Definition. *Frontiers in Psychology*, 13, 805795. <https://doi.org/10.3389/fpsyg.2022.805795>
- Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional Belief. *Annual Review of Psychology*, 62(1), pp. 271-298. <https://doi.org/10.1146/annurev.psych.121208.131622>
- Conrad, K. (1997). *La esquizofrenia incipiente: Ensayo de un análisis gestáltico del delirio* (1a. ed). Fundación Archivos de Neurobiología.
- Currie, G. (2000). Imagination, Delusion, and Hallucinations. *Mind & Language*, 15(1) , pp. 168-183. <https://doi.org/10.1111/1468-0017.00128>

Darwin, C. (1987). *El Origen de las Especies*. Espasa Calpe. <https://www.buscalibre.cl/libro-el-origen-de-las-especies/9788423918164/p/7641695>

Davies, M., & Coltheart, M. (2023). The two-factor theory of delusion. En E. Sullivan-Bissett (Ed.), *Routledge Handbook of the Philosophy of Delusion*. Routledge.

De Haan, S. (2020). An Enactive Approach to Psychiatry. *Philosophy Psychiatry & Psychology*, 27(1), pp. 3-25. <https://doi.org/10.1353/ppp.2020.0001>

DeVylder, J. E. (2019). Fixity of thinking and the foundations of identity: An argument for the evolutionary adaptiveness of delusions. *Early Intervention in Psychiatry*, 13(3) , pp. 720-721. <https://doi.org/10.1111/eip.12732>

Di Paolo, E. A., Buhrmann, T., & Barandiaran, X. E. (2017). *Sensorimotor life: An enactive proposal* (First edition). Oxford University Press.

Egan, A. (2008). Imagination, Delusion, and Self-Deception. En T. Bayne & J. Fernández (Eds.), *Delusions, Self-Deception: Affective and Motivational Influences on Belief Formation* (pp. 263-280). Psychology Press.

Feyaerts, J., & Sass, L. (2024). Self-Disorder in Schizophrenia: A Revised View (1. Comprehensive Review– Dualities of Self- and World-Experience). *Schizophrenia Bulletin*, 50(2) pp. 460-471. <https://doi.org/10.1093/schbul/sbad169>

Fineberg, S. K., & Corlett, P. R. (2016). The doxastic shear pin: Delusions as errors of learning and memory. *Cognitive Neuropsychiatry*, 21(1) , pp. 73-89. <https://doi.org/10.1080/13546805.2015.1136206>

Fuchs, T. (2005). Delusional Mood and Delusional Perception – A Phenomenological Analysis. *Psychopathology*, 38(3), pp. 133-139. <https://doi.org/10.1159/000085843>

Fusar-Poli, P., Estradé, A., Stanghellini, G., Venables, J., Onwumere, J., Messas, G., Gilardi, L., Nelson, B., Patel, V., Bonoldi, I., Aragona, M., Cabrera, A., Rico, J., Hoque, A., Otaiku, J., Hunter, N., Tamellini,

M. G., Maschião, L. F., & Maj, M. (2022). The lived experience of psychosis: A bottom-up review co-written by experts by experience and academics. *World Psychiatry*, 21(2), pp. 168-188. <https://doi.org/10.1002/wps.20959>

Garson, J. (2021). The developmental plasticity challenge to Wakefield's view. En *Defining mental disorder: Jerome Wakefield and his critics* (pp. 335-351). The MIT Press. <https://doi.org/10.7551/mitpress/9949.003.0021>

Gayon, J. (2009). Mort ou persistance du darwinisme? Regard d'un épistémologue. *Comptes Rendus Palevol*, 8(2-3) pp. 321-340. <https://doi.org/10.1016/j.crpv.2008.11.004>

Green, H., Hauser, L., & Troyakov, V. (2018). Are delusions beliefs? A qualitative examination of the doxastic features of delusions. *Psychosis*, 10(4) pp. 319-328. <https://doi.org/10.1080/17522439.2018.1528298>

Guidano, V. F. (1987). *Complexity of the self: A developmental approach to psychopathology and therapy*. Guilford Press.

Guidano, V. F. (1991). *The self in process: Toward a post-rationalist cognitive therapy*. Guilford Press.

Hemsley, D. R., & Garety, P. A. (1986). The formation of maintenance of delusions: A Bayesian analysis. *The British Journal of Psychiatry: The Journal of Mental Science*, 149 , pp. 51-56. <https://doi.org/10.1192/bjp.149.1.51>

Jaspers, K. (1993). *Psicopatología general* (R. O. Saubidet, Trad.; 2. ed). Fondo de Cultura Económica.

Kingsbury, J. (2006). A Proper Understanding of Millikan. *Acta Analytica*, 21(40), pp. 23-40. <https://doi.org/10.1007/s12136-006-1008-8>

Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanesh, S. J., Gur, R. E., & Gur, R. C. (2003). Facial emotion recognition in schizophrenia: Intensity effects and error pattern. *The American Journal of Psychiatry*, 160(10), pp. 1768-1774. <https://doi.org/10.1176/appi.ajp.160.10.1768>

Lancellotta, E. (2022a). Are delusions adaptive? *Wiley interdisciplinary reviews. Cognitive science*, 10(5), e1502. <https://doi.org/10.1002/wcs.1502>

Lancellotta, E. (2022b). Is the biological adaptiveness of delusions doomed? *Review of Philosophy and Psychology*, 13(1) pp. 47-63. <https://doi.org/10.1007/s13164-021-00545-6>

Lancellotta, E., & Bortolotti, L. (2019). Are clinical delusions adaptive? *WIREs Cognitive Science*, 10(5), e1502. <https://doi.org/10.1002/wcs.1502>

Lancellotta, E., & Bortolotti, L. (2020). Delusions in the two-factor theory: Pathological or adaptive? *European Journal of Analytic Philosophy*, 16(2) pp. 37-57. <https://doi.org/10.31820/ejap.16.2.2>

López-Silva, P. (2017). The typology problem and the doxastic approach to delusions. *Filosofía Unisinos*, 17(2), pp. 202-211.

López-Silva, P. (2014). La relevancia filosófica del estudio de la esquizofrenia. Cuestiones metodológicas y conceptuales. *Revista Colombiana de Psiquiatría*, 43(3), pp. 168-174. <https://doi.org/10.1016/j.rcp.2014.07.002>

López-Silva, P. (2018). Mapping the Psychotic Mind: A Review on the Subjective Structure of Thought Insertion. *Psychiatric Quarterly*, p. 89. <https://doi.org/10.1007/s11126-018-9593-4>

López-Silva, P. (2022). La marca de la psicosis: Hacia una síntesis del problema tipológico de los delirios. *Revista Colombiana de Psiquiatría*, 52, S183-S189. <https://doi.org/10.1016/j.rcp.2021.11.002>

López-Silva, P. (2023a). Creencias e imaginaciones: Reexaminando el problema tipológico de los delirios. *Pensamiento. Revista de Investigación e Información Filosófica*, 79(302), pp. 273-286. <https://doi.org/10.14422/pen.v79.i302.y2023.014>

López-Silva, P. (2023b). Minimal Biological Adaptiveness and the Phenomenology of Delusions in Schizophrenia. En A. Falcato & J. Gonçalves (Eds.), *The Philosophy and Psychology of Delusions* (1.a ed., pp. 126-140). Routledge. <https://doi.org/10.4324/9781003288992-10>

López-Silva, P., & Cavieres, A. (2022). El delirio en la esquizofrenia como objeto de estudio interdisciplinario en la filosofía de la mente. *Sophía*, 33 , pp. 71-90. <https://doi.org/10.17163/soph.n33.2022.02>

López-Silva, P., & Cavieres, A. (2023). Schizophrenia and the Error-Prediction Model of Thought Insertion. En P. López-Silva & T. McClelland (Eds.), *Intruders in the Mind: Interdisciplinary Perspectives on Thought Insertion* (pp. 113-134). Oxford University Press. <https://doi.org/10.1093/med/9780192896162.003.0008>

López-Silva, P., & Abarca, M. (2023). La hipótesis de la saliencia aberrante: Unificando la neurobiología y la fenomenología de la esquizofrenia. *Revista Latinoamericana de Psicopatología Fundamental*, 26, e220421. <https://doi.org/10.1590/1415-4714.e220421>

Maiese, M. (2021). An enactivist reconceptualization of the medical model. *Philosophical Psychology*, 34(7) , pp. 962-988. <https://doi.org/10.1080/09515089.2021.1940119>

McGinn, C. (2004). *Mindsight: Image, dream, meaning*. Harvard University Press.

McKay, R. (2012). Delusional Inference. *Mind and Language*, 27(3), pp. 330-355. <https://doi.org/10.1111/j.1468-0017.2012.01447.x>

McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioural and Brain Sciences*, 32(6) pp. 493-510. <https://doi.org/10.1017/S0140525X09990975>

Millikan, R. G. (1986). Thoughts Without Laws; Cognitive Science with Content. *The Philosophical Review*, 95(1) pp. 47-80. <https://doi.org/10.2307/2185132>

Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice*. MIT Press.

Motut, A., Isaac, C., Castillo, M.-C., & Januel, D. (2023). Link between metacognition and social cognition in schizophrenia: A systematic review and meta-analysis. *Frontiers in Psychiatry*, 14. <https://doi.org/10.3389/fpsy.2023.1285993>

- National Institute of Mental Health. (2015). *La esquizofrenia* (NIH). [https://www.nimh.nih.gov/sites/default/files/documents/health/publications/schizophrenia/schizophrenia\\_1.pdf](https://www.nimh.nih.gov/sites/default/files/documents/health/publications/schizophrenia/schizophrenia_1.pdf)
- Nesse, R. (1998). Emotional disorders in evolutionary perspective. *British Journal of Medical Psychology*, 71(4) pp. 397-415. <https://doi.org/10.1111/j.2044-8341.1998.tb01000.x>
- Nielsen, K. (2023). *Embodied, Embedded and Enactive Psychopathology: Reimagining Mental Disorder*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-29164-7>
- Parrott, M. (2021). Delusional Predictions and Explanations. *The British Journal for the Philosophy of Science*, 72(1) pp. 325-353. <https://doi.org/10.1093/bjps/axz003>
- Payne, R. (2013). *Speaking to My Madness: How I Searched for Myself in Schizophrenia*. CreateSpace Independent Publishing Platform.
- Pigliucci, M., & Müller, G. (Eds.). (2010). *Evolution, the extended synthesis*. MIT Press.
- Raup, D. M. (1992). *Extinction: Bad genes or bad luck?* (1. publ). Norton.
- Sass, L., & Feyaerts, J. (2024). Self-Disorder in Schizophrenia: A Revised View (2. Theoretical Revision—Hyperreflexivity). *Schizophrenia Bulletin*, 7(50) pp. 472-483. <https://doi.org/10.1093/schbul/sbad170>
- Schwitzgebel, E. (2012). Mad Belief? *Neuroethics*, 5(1), pp. 13-17. <https://doi.org/10.1007/s12152-011-9127-3>
- Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The Predictive Coding Account of Psychosis. *Biological Psychiatry*, 84(9) , pp. 634-643. <https://doi.org/10.1016/j.biopsych.2018.05.015>
- Sterzer, P., Mishara, A. L., Voss, M., & Heinz, A. (2016). Thought Insertion as a Self-Disturbance: An Integration of Predictive Coding and Phenomenological Approaches. *Frontiers in Human Neuroscience*, 10. <https://doi.org/10.3389/fnhum.2016.00502>

Van Duppen, Z. (2017). The Intersubjective Dimension of Schizophrenia. *Philosophy Psychiatry & Psychology*, 24(4), pp. 399-418. <https://doi.org/10.1353/ppp.2017.0058>



# *Hypotheses fingo*: interpretación, anticipación e investigación de la naturaleza en los *Principia Mathematica* de Newton

Andrew Mayne-Nicholls Merino<sup>a</sup>

## Resumen

En este artículo se intentará a) comprender el método de descubrimiento de Newton dentro del contexto de (i) la disputa entre las dos tradiciones preponderantes en la *Royal Society* – la tradición de los naturalistas y la de los matemáticos – y (ii) la influencia de la propuesta de Francis Bacon sobre la corriente naturalista; para luego b) problematizar la famosa afirmación de Newton *hypotheses non fingo*, examinando qué entiende Newton por *hypothesis* y en qué medida dicha afirmación es consistente con el resto de la propuesta presente en los *Principia*; y, finalmente, c) proponer una cuarta interpretación —sumada a las ya presentadas por Andrew Janiak— según la cual se podría comprender a Newton como un baconiano moderado.

**Palabras clave:** Hipótesis, Anticipación, Interpretación, Naturalistas, Matemáticos.

## Abstract

This paper will attempt to a) understand Newton's method of discovery in the context of (i) the dispute between the two prevailing traditions in the Royal Society - the naturalist tradition and the mathematician tradition - and (ii) the influence of Francis Bacon's proposal

---

<sup>a</sup>Estudiante de Licenciatura en Filosofía, Universidad de Chile.  
Contacto: a.maynenicholls@ug.uchile.cl

on the naturalist current; and then b) to problematize Newton's famous statement *hypotheses non fingo*, examining what Newton means by *hypothesis* and to what extent this statement is consistent with the rest of the proposal in the *Principia*; and, finally, c) to propose a fourth interpretation - in addition to those already presented by Andrew Janiak - according to which Newton could be understood as a moderate Baconian.

**Keywords:** Hypothesis, Anticipation, Interpretation, Naturalists, Mathematicians.

## 1. Introducción

Como bien muestra Steffen Ducheyne, Newton traza una serie de críticas epistemológicas contra Leibniz y Descartes, según las cuales desde una perspectiva empirista se niega a aceptar las ideas innatas de Descartes en la medida que son meras hipótesis, mientras que él se enfocaría solamente en hacer demostraciones matemáticas a partir de los fenómenos (Ducheyne 2006, p. 246 ss.). Ciertamente este enfoque permite explicar las razones que llevaron a Newton a escribir su connotado dictum: *hypotheses non fingo*.

Sin embargo, este no es el único conflicto que permite explicar el énfasis de Newton por aclarar que no inventa hipótesis, pues no es necesario salir de Inglaterra para encontrar otras tantas disputas y tensiones en las que se vio envuelto este autor. En efecto, en el seno de la *Royal Society of London*, que Newton presidiera por casi 25 años, podemos encontrar un debate similar en donde el bando al que él era más próximo (i.e. el bando de los matemáticos) recibió de parte del otro bando (i.e. los naturalistas) la misma crítica que el propio Newton lanzó contra Descartes y Leibniz, a saber: la de inventar hipótesis. En este sentido, el propósito de este artículo es analizar las propuestas metodológicas que Newton expresa en el *Escolio General* de sus *Principia Mathematica* desde la perspectiva de la disputa que se dio al interior de la *Royal Society*.

Como bien documenta Mordechai Feingold, la *Royal Society* de Newton estuvo marcada por la división interna entre naturalistas y matemáticos (Feingold 2001). De estos, los primeros apelaban, en concordancia con el espíritu de la inducción baconiana, a la recolección paciente y paulatina de datos mediante la observación y experimentación. Por el contrario, la tradición matematizante apelaba a la necesidad de aunar este confuso amasijo de información en teorías que cuenten con la certeza y precisión de la geometría y las matemáticas en general (Feingold 2001). El punto interesante que cabe destacar es el hecho de

que Newton, que se jactaba de no inventar hipótesis y de seguir un método de deducción a partir de los mismos fenómenos, era parte de la facción matemática. Esta facción recibía la constante acusación de fabricar hipótesis y fantasías por parte de los naturalistas, quienes defendían ante todo la experimentación y la observación de los fenómenos como una piedra angular del proceder científico. Un procedimiento que, en la línea de Francis Bacon, ameritaba un desarrollo inevitablemente lento para el posterior establecimiento de explicaciones científicas.

Es decir, en principio, Newton se encuentra en una posición al menos tensa, sino paradójica. Pues al mismo tiempo que niega rotundamente en el *Escolio General* de sus *Principia* que él elabore hipótesis, forma parte de aquella tradición<sup>1</sup> acusada de que “al amparo de la geometría y de otras amables especulaciones están estableciendo nuevas hipótesis” (Smith como se citó en Feingold 2001, p. 90).

A partir de esta tensión surge una serie de interrogantes: ¿qué entiende Newton por hipótesis? ¿Es fiel a su propia aserción de no fabricarlas? ¿En qué medida es compatible su método de descubrimiento científico –con sus idealizaciones, abstracciones y, más aún, conjeturas– con la propuesta baconiana?

## 2. Newton: un autor, dos tradiciones y tres interpretaciones

A partir de la controversia entre la tradición naturalista y la tradición matemática, re-presentantes respectivamente de la historia natural y de la física matematizada, se vuelve difícil comprender a Newton<sup>2</sup>. Desde la publicación de la primera edición de sus *Principia*, se ha escrito

---

<sup>1</sup>En este sentido Feingold afirma que: “La publicación de los *Principia* marcó una nueva fase en las relaciones entre los matemáticos y los naturalistas, no solo porque su éxito inmediato contribuyó en gran medida a la extendida aceptación de las pretensiones de la física matematizada (incluso por aquellos que no eran capaces de entenderla), sino que también porque el triunfo del método *newtoniano* parecía aprobar su aplicación a los otros dominios de la ciencia” (Feingold 2001, p.87). Esta traducción es propia. A menos que se indique lo contrario, también las traducciones siguientes serán de mi autoría. En cada caso se indicarán los detalles de la fuente.

<sup>2</sup>Por razones de tiempo y extensión este trabajo se enfocará principalmente en los *Principia*, particularmente, las Leyes, Reglas, Escolio y Escolio General; sin embargo, también se ilustrarán ciertos aspectos mediante pasajes de otras obras y de su correspondencia.

una ingente cantidad de páginas llenas de las más diversas interpretaciones (y malinterpretaciones) del autor.

En este contexto, Andrew Janiak presenta en su libro *Newton as a Philosopher* tres interpretaciones que buscan responder a dos preguntas claves: ¿cómo entender la afirmación *hypotheses non fingo*? Y, sobre todo: ¿cómo entender la relación entre física y metafísica dentro del corpus newtoniano? La primera interpretación propone entender a Newton como un agnóstico en metafísica; la segunda afirma que Newton es un empirista radical; mientras que la tercera, que es defendida por el propio Janiak, sostiene que Newton es un autor que invierte la jerarquía cartesiana entre física y metafísica.

La primera interpretación no es defendida, como el mismo Janiak aclara, por ningún autor en particular, pero podría considerarse la interpretación estándar que recibió Newton a partir del siglo XVIII en adelante (Janiak 2008, p.12). Desde esta perspectiva se lleva a cabo una interpretación bastante sofisticada de la afirmación *hypotheses non fingo*. Según ella, se debe entender esta declaración como un rechazo a la metafísica tradicional, representada principalmente por Descartes y Leibniz. Evidentemente, esta interpretación choca con el hecho de que Newton a menudo se inmiscuye en discusiones metafísicas, de las cuales la discusión sobre Dios, el espacio y el tiempo podrían ser los ejemplos por antonomasia (Janiak 2008, p. 29; cfr. Ducheyne 2006, p. 247 ss.).

Sin embargo, pese a que la evidencia textual muestra que Newton efectivamente se inmiscuyó en discusiones metafísicas, esta interpretación tendería a sortear este problema apelando a un “mero residuo histórico, uno que no juega ningún rol en su obra (*work*) principal en los *Principia*” (Janiak 2008, p.12). Sin embargo, no deja de ser problemático que esta interpretación obvia el hecho de que a lo largo de su obra Newton suele tomar parte activa en disputas metafísicas, y que incluso cuando Newton discute con Leibniz acerca de la causación, entra de hecho en una discusión de esta índole.

A partir de los problemas de esta interpretación surge una segunda lectura – propuesta en primer lugar por Stein y seguida a grandes rasgos por DiSalle – que propondría que Newton, lejos de evadir las preguntas metafísicas, lo que hace es responderlas sobre la base de una postura epistémica radicalmente empirista (Janiak 2008). De esta manera, Newton estaría transformando “la metafísica de una investigación *a priori* sobre la estructura del mundo en una investigación completamente empírica, que solo puede progresar mirando al desarrollo de la ciencia empírica” (Janiak 2008, p. 25). Es decir, abordaría las viejas

problemáticas sobre espacio, tiempo, movimiento, etc., pero con la diferencia de que ahora procedería a partir de los resultados seguros de la ciencia, otorgando una prioridad lógica a la física por sobre la metafísica.

Interpretar a Newton como un empirista radical no es para nada descabellado. En efecto, Ducheyne muestra, a partir de la lectura atenta de una serie de manuscritos, cómo varias – aunque ciertamente no todas – de las distancias que Newton guardó contra Descartes y Leibniz se deben a desacuerdos epistemológicos. Un claro ejemplo de esto es un pasaje de los manuscritos de Newton donde critica las ideas innatas de Descartes y afirma que incluso las ideas son parte de la física en la medida en que también ellas derivan de los fenómenos (CUL Add. Ms. 9597.2.11: f. 2r citado en Ducheyne 2006, pp. 22 y 23).

Frente a las ideas innatas que postula Descartes, Newton responde desde una perspectiva empirista que en realidad las ideas derivan meramente de los fenómenos empíricos y que por lo mismo su dilucidación concierne al estudio de la experiencia por parte de la física. Sin embargo, para Janiak, esta interpretación también choca con importantes obstáculos. Pues, si bien constituye un avance frente a la primera interpretación, se encuentra a la vez con el difícil asunto relativo a cómo interpretar la postura de Newton frente a ciertos problemas, como es el caso de Dios o de la acción a distancia.

Según Janiak, es un grave error suponer que Newton es un empirista en el mismo sentido que Bacon o Hume, como si intentara responder a la pregunta sobre cuál es la fuente del conocimiento humano. Vinculado con lo anterior, también supondría un error asumir que Newton tiene un carácter radicalmente empirista respecto a los temas metafísicos ya mencionados: tiempo, espacio, causación y, ciertamente, Dios. Pues por lo pronto queda la duda acerca de qué evidencia empírica podría servir a favor o en contra de la existencia de Dios, de sus atributos y de su carácter de necesario o contingente.

A partir de estos problemas, Janiak expone y defiende su propia interpretación. Según esta tercera lectura, que en parte se apoya en la anterior, Stein y Disalle aciertan al afirmar que Newton “transforma los asuntos de la metafísica mundana en preguntas empíricas en el mismo sentido de que nuestras respuestas a estas preguntas están siempre sujetas a refinamiento, revisión y rechazo sobre la base de la evidencia empírica” (Janiak 2008, p. 45). Sin embargo, el apoyo que da Janiak a Stein y DiSalle es limitado en la medida en que estos autores erran al

no distinguir entre lo que sería una metafísica divina y una metafísica mundana (Janiak 2008, p. 45 ss.).

Mientras que la metafísica divina versaría sobre temas vinculados a Dios, su naturaleza y su relación con el mundo, la metafísica mundana trataría sobre la naturaleza del movimiento, las fuerzas, la causalidad y otros fenómenos físicos. De esta forma, la primera de estas metafísicas constituye un marco básico y fundamental que engloba las investigaciones del mundo físico sin verse afectado por los resultados de dichas investigaciones. En cuanto que la metafísica mundana, por la naturaleza de los fenómenos que investiga, sí se ve interpelada por los resultados de las investigaciones empíricas (Janiak 2008, p.45).

De esta forma, la afirmación de Stein y DiSalle de que Newton coloca a la metafísica después de la física (en un orden de prioridad lógica) sería aplicable solamente a la metafísica mundana. Sin embargo, no ocurre lo mismo en el caso de la metafísica divina. Esta, lejos de estar sujeta a revisión a partir de los resultados seguros de la ciencia, constituye el marco (*framework*) de la investigación científica, marco que ordena la preferencia de ciertas interpretaciones por sobre otras y que incluso puede conducir a rechazar ciertas teorías:

Newton efectivamente transforma las cuestiones de metafísica mundana en cuestiones empíricas justo en el sentido de que nuestras respuestas a estas cuestiones están siempre sujetas a perfeccionamiento, revisión y rechazo sobre la base de la evidencia empírica. Pero esta investigación de la naturaleza se produce dentro de un marco fundamental que se fija independientemente de cualquier hallazgo empírico que hagamos (Janiak 2008, p.45).

La presente interpretación parece adecuada. Con todo, ella deja abierta la pregunta acerca de cómo interpretar la afirmación tan categórica de Newton: “Yo no fabrico (*feign*) hipótesis” (Newton 1999, p. 943). En este sentido, me parece que es posible proponer una cuarta interpretación que no busca ser contraria sino que más bien complementaría a la tercera. A partir de esta cuarta interpretación, podríamos comprender también a Newton y no solamente a los naturalistas como un heredero, con mayor o menor grado de fidelidad, de Bacon. Pero antes de llegar a este punto, vale la pena revisar la propuesta baconiana y su recepción en la *Royal Society of London*.

### 3. Bacon y la Royal Society

Como es bien sabido, Bacon busca en su *Novum Organon* la instauración de un nuevo método de descubrimiento. El fundador del empirismo moderno intenta reemplazar al modo de proceder que, en su opinión, se había llevado a cabo hasta su llegada, pues a su parecer antes se habría anticipado a la naturaleza en lugar de interpretarla a partir de la observación rigurosa y paulatina. Esta anticipación se debe en buena parte a que el ser humano caería en el error de confiar excesivamente en su entendimiento y en sus sentidos (Bacon 2000).

Así, nos encontramos en la filosofía de Bacon con una fuerte crítica a los alcances que tendría el entendimiento humano. Aquella precipitada tendencia de anticipar a la naturaleza que él tanto reprochaba a los escolásticos y a los antiguos, incluso se la puede ver en aquellos autores que Bacon reconoce como empiristas precedentes. Pues aquella tendencia, lejos de ser un vicio de algunas corrientes filosóficas, sería parte de “el prematuro y precipitado apuro de la mente y de su brincar o volar a las aserciones generales y los principios de las cosas” (Bacon 2000, p. 52). Más aún, fuera de esta tendencia puntual, para Bacon el entendimiento humano constituye un verdadero espejo irregular que distorsiona y confunde los rayos de luz que le llegan de la naturaleza (Bacon 2000, p. 41).

A su vez, Bacon muestra – pese a su marcado empirismo – una gran desconfianza respecto a las capacidades de los sentidos. Sin ir más lejos, él mismo admite que el método de anticipación de la naturaleza que él tanto denuesta – pues se precipitaría de las cosas particulares a los axiomas generales sin la certeza requerida – se construye sobre los datos de los sentidos. Estos últimos pueden inducir a error cuando no son dirigidos correctamente, en tanto no dan cuenta de la complejidad del universo. En este sentido, el filósofo inglés asegura que las percepciones no son ni de cerca la medida de las cosas (Bacon 2000, p. 41), cuya sutileza escapa a nuestros sentidos. Tanto así que este empirista incluso afirma que los engaños y las limitaciones de los sentidos constituirán uno de los obstáculos fundamentales del entendimiento humano:

Pero por lejos el mayor obstáculo y distorsión del entendimiento humano proviene de la torpeza, las limitaciones y los engaños de los sentidos; de modo que las cosas que golpean los sentidos tienen una mayor influencia que incluso las cosas más poderosas que no golpean directamente los sentidos (Bacon 2000, p. 45).

Es a partir de estas consideraciones que el autor afirma, como mencionábamos más arriba, que hasta su tiempo se había anticipado mas no interpretado a la naturaleza. De esta manera, lejos de desarrollar una metodología de investigación basada sobre la experiencia y sobre una observación rigurosa de los particulares, se ha saltado con demasiada ligereza desde pocas observaciones casuales a conclusiones apresuradas. Particularmente, los lógicos, en opinión de Bacon, han saltado rápidamente desde una mirada parcial de los particulares hacia axiomas generales. Sin embargo, a partir de aquí, los pasos intermedios entre uno y otro – e incluso la evidencia que se encuentre en contra de dichos axiomas – serán remediados por distinciones conceptuales y por axiomas medios que no hacen justicia a la sutileza de la naturaleza.

La interpretación de la naturaleza, por el contrario, consiste en llegar a los axiomas generales por una ascensión paulatina y progresiva. Así, mediante la sistematización de las observaciones, mediante tablas de inducción, se buscaría en primer lugar registrar la información pertinente a cierto fenómeno para luego aislar los factores relevantes y llegar a una conclusión más o menos segura (Bacon 2000).

De esta manera, considerando el espíritu marcadamente empirista e inductivista de Bacon, sumado a sus propuestas concretas para un nuevo método de descubrimiento, podemos comprender la importancia que tuvo esta figura en la conformación y consolidación de la *Royal Society*. En este sentido, no es casualidad el hecho de que autores como Lister, pertenecientes a la tradición naturalista, se jactaran de tener

también a mi señor (*lord*) Bacon de mi lado, quien fue el primer restaurador de esta clase de aprendizaje y [quien] dio el primer aviso (*notice*) de nuestros anhelos (*wants in*) y métodos sobre cómo desarrollar la Filosofía Natural sobre la verdad de los experimentos y observaciones y no [sobre] salvajes hipótesis matemáticas (Lister como se citó en Feingold 2001, p. 92).

Frente a la tradición matemática que fabricaría hipótesis que no están sustentadas por una observación rigurosa de los fenómenos naturales, la tradición naturalista ofrecería una base para erigir cualquier teoría posterior. Como afirma Feingold:

al menos desde la perspectiva de los naturalistas, los matemáticos socavaron la misma piedra angular sobre la cual se levantó la tradición del empirismo científico inglés: los experimentos y las observaciones (2001 p. 95).



De este modo, la experimentación y la observación constituyen el camino seguro por el que seguiría la investigación científica en el mismo espíritu que Bacon.

En efecto, Feingold afirma que si bien es cierto que “es debatible hasta qué punto la Sociedad adoptó la ideología baconiana” (Feingold 2001, p. 79) posturas como las de Thomas Sprat o Sir Robert Moray, quienes defendían la necesidad de desechar toda teoría o hipótesis incierta en virtud de la paciente observación y recolección de conocimiento seguro, no eran para nada minoritarias. Un claro testimonio de esta opinión lo podemos encontrar en el siguiente extracto de Moray citado por Feingold:

Mientras tanto, esta Sociedad no aceptará ninguna Hipótesis, sistema o doctrina de los principios de la Filosofía Natural propuestos o mantenidos por ningún Filósofo Antiguo o Moderno, ni la explicación de ningún fenómeno en el que se deba recurrir a causas originales (...) Tampoco definirá dogmáticamente ni fijará axiomas de asuntos científicos, sino que cuestionará y pondrá en tela de juicio todas las opiniones[,] no adoptando ni adhiriéndose a ninguna hasta que, mediante un debate maduro y argumentos claros, principalmente los que se deducen de experimentos legítimos, se demuestre invenciblemente la veracidad de tales posiciones. (Moray citado en Feingold 2001, p.79).

Lejos de inventar hipótesis o apresurarse a postular cualquier teoría, la tradición naturalista proponía limitarse al conocimiento cierto, aunque paulatino, producido por una observación y experimentación rigurosas. En este contexto, no sería forzado interpretar en términos baconianos la acusación que hacía la tradición naturalista de la siguiente forma: mientras que ella abriría el camino para una posterior interpretación de la naturaleza, la tradición matemática – en el natural apuro prematuro que tiene según Bacon el entendimiento humano – se precipitaría a fabricar anticipaciones que no son capaces de dar cuenta de la sutileza de la naturaleza.

Ahora bien, ciertamente este tipo de posicionamientos respecto al proceder científico defendido por la vertiente naturalista significó recibir varias críticas por parte del bando contrario. En efecto, la tradición matematizante acusó a la tradición naturalista de limitarse a coleccionar y catalogar plantas, insectos y conchas, dejando de lado la verdadera investigación científica. Feingold da un claro ejemplo de este tipo de

objeciones al citar la carta de Molyneux a Halley, fechada el 08 de abril de 1686:

surgió en la Sociedad un partido que rechazaba todo tipo de conocimiento útil, excepto la clasificación y archivo de conchas, insectos, peces, pájaros, etc. en sus distintas especies y clases; y a esto lo llamaron Historia Natural e Investigación de la Naturaleza (Molyneux citado en Feingold 2001, p. 85).

Sin embargo, lo curioso es que siempre y cuando se le llame historia natural y no filosofía natural, esta manera de investigar concuerda perfectamente con el esquema epistemológico propuesto por Bacon. Este esquema tiene una forma piramidal donde nos encontramos con que a la base de la metafísica y la física se encuentra la historia natural (Jalobeanu 2016, p. 51). Pero más aún, como muestra Silvia Manzo, Bacon, concibiendo a la historia como un espejo o un registro imparcial de las cosas, libre de cualquier corrupción por parte del entendimiento humano, entiende que la tarea de la historia natural consiste en hacer un reporte de los “hechos desnudos” (*naked facts*) de la manera más simple e imparcial posible sin agregar ningún tipo de juicio (Manzo 2012, p. 37).

Pero esto es exactamente por lo que abogaban naturalistas como Sprat, Moray o Lister. Desde este punto de vista, podemos considerar que la crítica que hacía Molyneux a los naturalistas no es del todo justa. Podría ser debatible si acaso es pertinente la objeción de Molyneux de que los naturalistas no llevan a cabo ninguna investigación de utilidad. Sin embargo, no parece ser justa la objeción de que llamen a ese quehacer historia natural, pues desde una perspectiva baconiana, la historia natural consiste justamente en eso:

Según esta división epistémica del trabajo, la historia natural proporciona los datos individuales seleccionados de acuerdo con los temas prescritos a cada historia y las instancias prerrogativas. Los axiomas de la filosofía natural o proposiciones generales deben inducirse a través de proposiciones que ascienden a partir de los datos en orden de generalidad creciente. Así, las generalizaciones inductivas están permitidas una vez que se dispone de la base histórica necesaria (Manzo 2012, p. 46).

De esta manera, los naturalistas no estarían olvidando la investigación pertinente de la filosofía natural, sino simplemente posponiéndola – de

acuerdo con el esquema baconiano – hasta el momento en que haya suficiente material o data coleccionado a partir de los experimentos y la observación.

Curiosamente, ya en este punto encontramos un primer paralelismo entre Bacon y Newton. Pues como afirmara James Jurin:

Aquel gran hombre [Newton] era consciente de que para ser filósofo hacía falta algo más que conocer el nombre, la forma y las cualidades evidentes de un insecto, un guijarro, una planta o una concha (...) Todos recordamos aquella frase que pronunciaba con tanta frecuencia, que la historia natural puede proporcionar materiales para la filosofía natural, pero no es filosofía natural (Jurin 1727 como se citó en Feingold 2001, pp. 77-78).

Si bien Jurin se apoya en la autoridad de Newton justamente para utilizarlo en perjuicio de los naturalistas, esta apelación es de doble filo. Ya que al mismo tiempo que enfatiza la importancia de la filosofía natural y muestra que no basta con la mera historia natural, Newton reconoce su adhesión a la división de las labores de investigación postulada por Bacon y seguida por los naturalistas. En este sentido, podemos ver que Newton, por lo menos en este punto particular, sigue los lineamientos baconianos. Pero antes de seguir examinando las semejanzas y los paralelos entre uno y otro autor, es necesario examinar el *dictum* metodológico de Newton y revisar en qué medida fue en la práctica fiel a él.

#### 4. El curioso caso de las *hypotheses* de Newton

Como se dijo en un comienzo, encontramos en la obra de Newton una particular tensión. Por un lado, este filósofo se jacta – contra Descartes y Leibniz – de no inventar hipótesis, pero por otro lado se sitúa al interior de la *Royal Society* en el bando matemático, el cual recibe la constante acusación de estar inventando hipótesis y conjeturas. Pues desde el punto de vista de los naturalistas se podría afirmar, en jerga baconiana, que Newton anticipa la naturaleza en vez de interpretarla a partir de conocimiento rigurosamente recolectado. O se podría afirmar, como hacía Smith, citado anteriormente, que Newton se estaría amparando en la geometría para establecer meras hipótesis. Sin embargo, ¿es ese el caso? ¿estaría Newton anticipando o interpretando la naturaleza?

Es difícil encajar a Newton en cualquiera de estos modos de investigación y determinar si acaso anticipa (mediante hipótesis y especulaciones) o interpreta (mediante conocimiento cierto) a la naturaleza. Ya que, si bien realiza las idealizaciones y abstracciones de la tradición matemática, sería bastante atrevido acusarlo de no dar cuenta de la sutileza de la naturaleza dado el nivel de predictibilidad que otorgaron sus teorías y la unificación de los fenómenos supra y sub lunares que entrañaron. Para examinar con justicia si acaso Newton inventa hipótesis, conviene ante todo partir de la definición que el propio Newton da de este término:

Lo que sea que no es deducido a partir de los fenómenos debe ser llamado una hipótesis; y las hipótesis, ya sean metafísicas o físicas o basadas en cualidades ocultas o mecánicas, no tienen cabida en la filosofía experimental. En esta filosofía experimental [i] las proposiciones son deducidas a partir de los fenómenos y [ii] se vuelven generales mediante la inducción (Newton 1999, p. 943).

Podemos decir, por tanto, que para que un enunciado con pretensiones de verdad pueda ser considerado una hipótesis deberá cumplir con: (i) no haber sido deducido a partir de los fenómenos y (ii) haber sido generalizada mediante inducción. Como bien señala Cohen en su artículo de 1969, esta actitud de Newton respecto a las hipótesis no fue siempre igual de estricta. En efecto, considera un error asumir de forma axiomática que Newton no haya cambiado en este respecto lo más mínimo a lo largo de una trayectoria tan extensa (Cohen 1969, p. 304).

Según Cohen, Newton habría tenido dos posturas respecto a las hipótesis: el primer Newton se habría mostrado abierto a aceptar hipótesis siempre que se limiten a (a) explicar teorías o propiedades de las cosas o (b) sugerir futuras investigaciones, mientras que el segundo Newton habría pasado a ser mucho más crítico de la noción de hipótesis. De esta forma, en su obra más temprana (entre 1685 y 1692), Newton habría estado dispuesto a aceptar que “las hipótesis que buscan explicar teorías tienen su lugar en la ciencia, pero sólo como explicaciones, no prejuzgando la investigación sino haciéndola avanzar e incluso sugiriendo experimentos” (Cohen 1969, p. 322). Pero posteriormente vemos que hay un gran vuelco en su concepción de las hipótesis, ya que adoptó una gran aversión hacia ellas, negándose a aceptar que “los fundamentos de su Sistema del Mundo (en el Libro III de los *Princi-*

*pia*) podrían ser hipótesis en cualquiera de los sentidos corrientes de la palabra o que las conclusiones de la *Óptica* podrían ser un conjunto de hipótesis” (Cohen 1969, p. 322).

Ahora bien, pese a lo que dijo el propio Newton al final de su carrera, podemos mostrarnos escépticos frente a la afirmación de que él no haya postulado hipótesis. Por lo pronto, en el mismo *Escolio General* en donde Newton escribe este famoso *dictum* metodológico, podemos encontrar tres afirmaciones que podríamos considerar como hipótesis. Pues bajo su propia definición del término, realiza tres declaraciones que en estricto rigor no se pueden deducir a partir de los fenómenos (infringiendo así el primer requisito) ni fueron generalizadas por inducción (infringiendo así el segundo requisito).

La primera de estas afirmaciones hace referencia a la estructura del sistema solar, del cual Newton afirma que en vistas de su elegancia y complejidad no podría “haber surgido sin el diseño y dominio de un ser inteligente y poderoso” (Newton 1999, p. 940). Otorguemos que la existencia de Dios podría ser deducida a partir de los fenómenos<sup>3</sup>, mediante la versión del argumento cosmológico que Newton esgrime. Con todo, aun cuando la afirmación de existencia de Dios no sería una hipótesis (asumiendo que Newton toma esto a partir del conocimiento teológico), la afirmación de que la estructura del sistema solar no pudo sino ser diseñada por aquel ente excelentísimo ciertamente lo es. Y lo es bajo los propios parámetros de Newton<sup>4</sup>. En efecto, el hecho de que Kant y Laplace hayan podido ofrecer una explicación convincente del fenómeno sin apelar a Dios es prueba de que la afirmación de Newton (i) no es deducida a partir de los fenómenos y que (ii) tampoco es generalizada mediante inducción.

---

<sup>3</sup>Es debatible si acaso podemos deducir, de manera indirecta, la existencia de Dios. Sin embargo, el mismo Newton afirma que: “no hay ninguna sensación (*sense*) directa y no hay acciones indirectas reflejadas por las cuales conocemos las sustancias íntimas; mucho menos tenemos una idea de la sustancia de Dios. Solo lo conocemos mediante sus propiedades y atributos, y por la más sabia y óptima construcción de las cosas y de sus causas finales, y lo admiramos por sus perfecciones; pero lo veneramos y adoramos por su dominio” (Newton 1999, p. 942).

<sup>4</sup>Ciertamente, para Newton “tratar de Dios a partir de los fenómenos ciertamente forma parte de la filosofía ‘natural’” (Newton 1999, p. 943), por lo que esta objeción podría parecer un anacronismo. Sin embargo, del hecho de que podamos investigar a Dios a partir de los fenómenos no se sigue que sea legítimo concluir que la estructura del Sistema Solar solo se puede explicar apelando a su intervención.

La segunda hipótesis también trata sobre Dios. En la misma línea de la afirmación anterior, Newton señala que “es de común acuerdo que el supremo Dios existe necesariamente y que por la misma necesidad él es (*is*) siempre y en todo lugar” (Newton 1999, p. 942). Al igual que en el caso anterior, aun concediendo que se pueda deducir indirectamente a partir de los fenómenos la afirmación de que Dios existe, no es posible deducir ni generalizar mediante inducción el hecho de que Dios: (i) sea necesario y (ii) sea eterno. Una situación contrafáctica - por ejemplo, que la existencia de Dios sea, después de todo, contingente y que el día de mañana Dios deje de existir - podría perfectamente darse al mismo tiempo que los fenómenos naturales a los que apela Newton.

La tercera hipótesis, que da varias luces sobre cómo entender el rechazo a la fabricación de hipótesis por parte de Newton, es la especulación sobre cierto “espíritu sutilísimo [que] impregna los cuerpos crasos (*gross bodies*) y yace oculto en ellos; por su fuerza y acciones, los cuerpos particulares se atraen unos a otros a distancias muy pequeñas y se cohesionan cuando llegan a ser contiguos” (Newton 1999, p. 944)<sup>5</sup>. Incluso para el mismo Newton, esta afirmación no es posible (i) deducirla a partir de los fenómenos ni (ii) generalizarla a partir de la inducción<sup>6</sup>.

Ahora bien, ¿cómo se podría explicar que en la misma sección de los *Principia* Newton declare explícitamente que él no inventa hipótesis y, acto seguido, que fabrique tres? Me parece que la clave está en esta tercera hipótesis, pues ella da cuenta de la intención epistémica que tiene Newton. Como bien señala Janiak, aunque muchas veces se pasa por alto, no hay que perder de vista lo sutil que es Newton respecto a las hipótesis (Janiak 2008, p. 19 y ss.). Una cosa es conjeturar o especular una posible explicación de un fenómeno – explicación que podría guiar y fomentar una futura investigación – y otra muy distinta es dar dicha hipótesis – al estilo de Descartes o Leibniz – como una explicación satisfactoria<sup>7</sup>

<sup>5</sup>Un ejemplo parecido puede encontrarse en la apelación al éter en la carta de Newton a Boyle del 28 de febrero de 1679.

<sup>6</sup>El mismo Newton señala que: “no existe un número suficiente de experimentos para determinar y demostrar con precisión las leyes que gobiernan las acciones de este espíritu” (Newton 1999, p. 944)

<sup>7</sup>Un gran ejemplo del rechazo de las hipótesis cartesianas está en la crítica que hace Newton de la teoría de Descartes sobre la naturaleza de los cometas (los que, según Descartes, serían soles que pierden su luz), en la Primera Carta a Bentley. Newton

Pero esa misma proposición puede ser considerada como un incentivo (*prod*) para una investigación empírica posterior; no es “fabricada” (*feigned*) a menos que uno adopte una injustificada actitud epistémica hacia ella, por ejemplo, si uno asevera que ella es la explicación correcta de algún fenómeno natural documentado razonablemente bien (Janiak 2008, p.19, las cursivas son mías).

De esta forma, vemos que la sutil actitud que Janiak observa en Newton es equivalente a la postura del primer Newton del que habla Cohen. Así, siguiendo a Janiak, podríamos ver un elemento de continuidad entre los dos Newtons de Cohen. Pues aun cuando al final de su obra Newton afirmara no realizar ningún tipo de hipótesis, lo cierto es que mantendría una misma actitud epistémica respecto a las conjeturas a lo largo de toda su carrera.

Me parece que esta actitud se manifiesta claramente en la carta de Newton a Boyle del 28 de febrero de 1679 – es decir, 34 años antes de la segunda edición de los *Principia* y, por tanto, de su aseveración de no fabricar hipótesis – donde afirma: “apenas puedo estimar digno de ser comunicado a otros aquello con lo cual no estoy satisfecho, sobre todo en la filosofía natural, *donde el fantasear no tiene fin*” (Newton 2004, p.1, las cursivas son mías). En este sentido, si bien durante los años cambió lo que Newton llamó hipótesis, se mantuvo constante su rechazo a cualquier explicación meramente especulativa que intente reemplazar la investigación empírica de la naturaleza.

Pues como afirma Janiak: “Newton era muy consciente de los límites del conocimiento de la naturaleza, límites que él consideraba que sus interlocutores y críticos transgredían al proponer hipótesis” (Janiak 2004, pp. xxx - xxxi). El ejemplo por antonomasia es el de los vórtices cartesianos, hipótesis que para Newton, como él mismo afirma en el comienzo de su *Escolio General* “tropieza con muchas dificultades” (Newton 1999, p. 939). Sin embargo, en la práctica Newton no rechaza las hipótesis que suponen un estímulo para la investigación.

De esta forma, a las condiciones para considerar una afirmación como hipótesis – a saber: (i) no ser deducido de los fenómenos y (ii) no

---

rechaza esta hipótesis pues va en contra de los fenómenos que él ha observado (los cometas van y vuelven) por lo que afirma que esta hipótesis “no tiene cabida en mi sistema y es manifiestamente errónea” (Newton 1998, p. 152).

ser generalizados mediante inducción – cabe agregar una tercera: (iii) tener la actitud epistémica de aseverar la verdad de dicha afirmación.

## 5. Una cuarta interpretación: Newton como un baconiano moderado

En la sección anterior se examinaron tres afirmaciones que podrían ser catalogadas como hipótesis. La primera de estas, referente a la intervención divina en la conformación del sistema solar – en la medida en que no fomenta sino que reemplaza y trunca cualquier futura investigación – no sería admisible ni siquiera por el primer Newton según los criterios impuestos por él mismo. Sin embargo, las otras dos sí son admisibles. Pues de ellas, la primera, aquella sobre los atributos divinos, conformaría según Janiak un marco metafísico divino que ordena las diferentes investigaciones empíricas. Mientras que la última hipótesis, respecto a aquel espíritu sutilísimo, correspondería a lo que Cohen llama una *working hypothesis* (Cohen 1969, p. 322), es decir, una hipótesis que estimula futuras investigaciones<sup>8</sup>.

Sin embargo, un caso mucho más complejo es el que se expone en la sección de las llamadas *Regulae Philosophandi*. Pues aquí nos encontramos con una serie de sentencias que ni son deducidas a partir de los fenómenos ni constituyen el marco metafísico divino del que habla Janiak. El ejemplo más claro es la Regla III donde Newton afirma que:

las cualidades de los cuerpos solo pueden ser conocidas mediante experimentos; y por lo tanto las cualidades que cuadran universalmente con experimentos deben ser consideradas como cualidades universales (...). Ciertamente no deben (*ought not*) fabricarse imprudentemente fantasías ociosas contra la evidencia de los experimentos, ni debemos (*should*) apartarnos de la analogía de la naturaleza, *puesto que la naturaleza es siempre simple y siempre concordante consigo misma* (Newton 1999, p. 795, las cursivas son mías).

Ciertamente acá no nos encontramos con una afirmación empírica. Por lo pronto, para saber que la universalidad de las cosas efectivamente

---

<sup>8</sup>Se podría considerar que las *Cuestiones (Queries)* de la *Óptica* pertenecen a este mismo grupo, aun cuando al final de su obra el propio Newton no lo admitiría (cfr. Cohen 1969; Janiak 2004, pp. 127 ss.; Torretti 1998, pp. 154 ss.).



concuenda con la universalidad de nuestros experimentos, habría que realizar experimentos sobre el universo entero. Experiencia que es evidentemente imposible. Y si por el contrario se intentara demostrar la concordancia de la universalidad de las cosas a partir de la universalidad de nuestros experimentos, se estaría apelando justamente a aquello que se intenta probar.

Sea como fuere, este tipo de afirmaciones – pese a lo que se podría pensar – no aleja a Newton de Bacon, sino que por el contrario traza un fuerte paralelismo entre ambos. Paralelismo que se puede evidenciar en dos casos en los que estos filósofos realizan una forma peculiar de deducción a partir de los fenómenos. El primer caso se encuentra en el artículo de Manzo anteriormente citado, donde la autora señala que:

Muchos fenómenos naturales imperceptibles se infieren a partir de hechos observados directamente en ensayos experimentales. Así, por ejemplo, Bacon suponía a veces la existencia de corpúsculos materiales y espacios vacíos a partir de la solución perceptible de azafrán en agua. Aunque no eran visibles, su existencia se suponía a partir de ciertos hechos perceptibles (Manzo 2012, p. 45).

En tanto que en el caso de Newton nos encontramos con el siguiente pasaje citado en el artículo de Ducheyne:

Hasta ahora he explicado el sistema de este mundo visible en lo que concierne a los movimientos mayores que pueden ser fácilmente detectados. Hay, sin embargo, innumerables movimientos locales que debido a la minuciosidad de las partículas en movimiento no pueden ser detectados (...) Puedo decir brevemente, sin embargo, que la naturaleza es extremadamente simple y conforme a sí misma. Cualquier razonamiento que se aplique a los movimientos más grandes debería aplicarse también a los más pequeños (Newton citado en Ducheyne 2006, p. 230).

Llama la atención que ambos autores realizan en cada caso el mismo tipo de inferencia, la que siguiendo a Ducheyne podríamos llamar transducción (Ducheyne 2006, cfr. p. 230 ss.), según la cual a partir de la observación de ciertas propiedades se puede asumir una correspondencia entre estas propiedades perceptibles y otras imperceptibles. De esta forma, la noción de deducción a partir de los fenómenos se ensancha

abarcando no solamente la generalización por inducción, sino también la transducción por analogía.

A mi parecer, este ejemplo muestra cómo no basta con la mera observación o recolección de los fenómenos. En términos baconianos, limitarse a aquello correspondería a hacer solamente historia natural, pero para realizar filosofía natural propiamente tal, habría que ir más allá e interpretar la naturaleza. Más allá de la agrupación y clasificación de datos particulares, es necesario deducir algo a partir de esos datos, obligando al filósofo natural a dar un salto interpretativo (cfr. Bacon 2000, pp. 51, 79, 81, *et passim*). Sin embargo, para que este salto se realice con rigurosidad y no caiga en las anticipaciones tan denostadas por Bacon, es necesario contar con un marco.

Siguiendo el análisis de Janiak, podríamos afirmar que en el caso de Newton este marco corresponde al marco dado por la metafísica divina que orienta las investigaciones sin verse afectada por sus resultados. En el caso de reglas para la investigación que parten de la base de determinada concepción de la naturaleza – como es el caso de la *Regla III* citada más arriba – me parece que dentro del esquema de Janiak habría que englobarlas no en el marco metafísico divino, sino en el mundano. Puesto que estas reglas conforman un marco que no trata sobre Dios ni su naturaleza, sino simplemente sobre el mundo físico y guían la investigación sin que hayan sido previamente deducidas a partir de los fenómenos. Con la única salvedad, por cierto, que estas reglas nunca se ven afectadas por los resultados de las investigaciones, cosa que Janiak no admitiría para el resto del marco metafísico mundano.

Habiendo dicho todo esto, me parece que podemos postular una cuarta interpretación según la cual Newton sería un baconiano moderado en el sentido de que investiga la naturaleza siguiendo las exigencias metodológicas de Bacon y, de esta manera, sorteando las objeciones de los naturalistas. En primer lugar, sería baconiano en la medida en que busca seguir un método de descubrimiento afín a la práctica de “interpretación” de la naturaleza que defiende Bacon: en efecto, Newton busca (i) deducir a partir de los fenómenos, (ii) generalizar mediante una inducción y (iii) no aseverar ningún conocimiento incierto con una actitud epistémica indebida. Por (i) cumple con el requerimiento baconiano de observar y experimentar. Por (ii) cumple con la precaución de no caer en lo que según Bacon sería la premura natural al entendimiento humano; lo que a su vez lleva a Newton a cumplir con (iii), al no caer en la errada actitud epistémica de reemplazar una investigación por meras “anticipaciones” de la naturaleza.

Por otro lado, Newton sería moderado no porque caiga en ciertas anticipaciones, sino porque cambia ligeramente el marco que guía las investigaciones baconianas. Así, frente a la naturaleza sutil de Bacon, Newton propone una naturaleza simple y concordante consigo misma.

## 6. Conclusión

En este artículo se buscó problematizar, dentro del contexto de los debates al interior de la *Royal Society*, la afirmación newtoniana de *hypotheses non fingo* y su relación con las hipótesis en general. A su vez, se examinaron las críticas que el bando de los matemáticos, al que pertenecía Newton, recibía de parte de los naturalistas, quienes les acusaban de postular hipótesis y conjeturas. A partir de un breve análisis de la postura naturalista y la influencia que en ella ejerció la figura de Francis Bacon, se revisó si acaso la obra de Newton era vulnerable a la crítica de los naturalistas por postular hipótesis o, en la jerga baconiana, anticipar la naturaleza.

Luego se intentó mostrar de qué manera parecería que Newton (bajo sus propios criterios) sí realiza hipótesis, pero que sin embargo tiene una actitud muy peculiar y sutil frente a ellas. Pues en efecto, atendiendo a la actitud epistémica con la que se aproxima a dichas hipótesis, podemos afirmar que él, en el esquema de Bacon, no anticipa la naturaleza (salvo en el caso de la primera de las hipótesis analizadas), sino que se limita a interpretarla.

De esta manera, a partir de todo lo anterior se propuso que Newton de hecho sí era consistente con los criterios metodológicos baconianos, pero de forma moderada. De esta manera, su método de investigación lograría abordar geoméricamente la naturaleza sin quedar expuesto a las objeciones del bando naturalista<sup>9</sup>.

## Referencias

Bacon, F. (2000). *Novum Organon*. Cambridge University Press.

Ducheyne, S. (2006). The General Scholium: Some Notes on Newton's Published and Unpublished Endeavours. *Lias: Sources and Documents Relating to the Early Modern History of Ideas*, 33(3), pp. 223-274.

---

<sup>9</sup>Agradezco al profesor Cristián Soto por los valiosos comentarios que hizo a una versión anterior de este artículo.

Feingold, M. (2001). *Mathematicians and Naturalists: Sir Isaac Newton and the Royal Society*. En Buchwald y Cohen (Eds.), *Isaac Newton's Natural Philosophy* (pp. 77-102). The MIT Press.

Janiak, A. (Ed.). (2004). *Newton: Philosophical Writings*. Cambridge University Press.

Janiak, A. (2008). *Newton as Philosopher*. Cambridge University Press.

Jalobeanu, D. (2016). "The Marriage Of Physics With Mathematics": Francis Bacon on Measurement, Mathematics, and the Construction of a Mathematical Physics. In G. Gorham, B. Hill, E. Slowik, & C. K. Waters (Eds.), *The Language of Nature: Reassessing the Mathematization of Natural Philosophy in the Seventeenth Century* (pp. 51-80). University of Minnesota Press. <https://doi.org/10.5749/j.ctt1d390rg.5>

Manzo, S. (2012). Francis Bacon's Natural History and Civil History: A Comparative Survey. *Early Science and Medicine*, 17(1/2) pp. 32-61. <http://www.jstor.org/stable/41723181>

Newton, I. (1999). *The Principia: Mathematical Principles of Natural Philosophy*. (Cohen et al., Trans.). University of California Press.

Torretti, R. (1998). *Filosofía de la Naturaleza*. Editorial Universitaria.

# Formas y pragmatismo en la filosofía de Francis Bacon

Sofía Muñoz Nilo<sup>a</sup>

## Resumen

El presente ensayo analiza la relación entre dos aspectos de la propuesta del filósofo Francis Bacon en su obra *Novum Organum* (NO): la visión pragmática de la ciencia y el conocimiento de formas al que el conocimiento humano aspira en general. Este análisis se desarrolla en consideración de distintas interpretaciones del concepto baconiano de forma a la luz del empirismo y pragmatismo del autor. Entre estas interpretaciones algunas han sido preferidas por su adecuación y por su contribución a la comprensión de la relación revisada permitiendo vislumbrar cómo la ciencia, el conocimiento de formas y la práctica se articulan para mejorar la situación humana.

**Palabras clave:** naturaleza simple, estructura corpuscular, leyes, reglas.

## Abstract

This essay analyses the relationship between two aspects of the philosopher Francis Bacon's proposal in his work *Novum Organum* (NO): the pragmatic vision of science and the knowledge of forms to which human knowledge aspires in general. This analysis is developed in consideration of different interpretations of the Baconian concept of form, in the light of the author's empiricism and pragmatism. Among these interpretations, some have been preferred for their appropriateness and for their contribution to the understanding of the relationship under review, allowing a glimpse of how science, knowledge of forms and practice are articulated to improve the human situation.

**Keywords:** simple nature, corpuscular structure, laws, rules.

---

<sup>a</sup>Universidad de Chile, Santiago, Chile.  
Contacto: sofia.munoz.nilo@ug.uchile.cl

## 1. Introducción

Francis Bacon ha sido considerado como una de las principales figuras de la filosofía natural del periodo de transición entre el Renacimiento y los principios de la edad moderna (Klein 2020). Entre sus principales propuestas se encuentran (i) su empirismo experimental, de modo que “la mejor demostración es con mucho la experiencia, siempre que vaya unida al experimento directo” (NO I 70)<sup>1</sup>; (ii) su concepción del conocimiento como vinculado estrechamente a la acción humana, siendo “la meta verdadera y legítima de las ciencias (...) dotar a la vida humana de nuevos inventos y recursos” (NO I 81); y (iii) su distanciamiento de la filosofía natural especulativa y desligada de la experiencia (NO I 62).

Estas posiciones se reflejaron en su esfuerzo por establecer la observación concreta como base para la comprensión de la naturaleza en su especificidad, partiendo siempre desde hechos particulares, avanzando hacia los principios generales de manera progresiva y evitando generalizaciones apresuradas; y en su insistencia en el rol fundamental del conocimiento para la dirección de la actividad humana, de modo que lo que en la observación es como causa es como regla en la práctica (NO I 3), permitiendo la modificación eficiente del mundo.

Por estos rasgos de su filosofía, sorprende que en su obra aparezcan términos como ‘forma’, muy utilizado por tradiciones que él mismo consideró como filosofías que “se asienta(n) sobre una base demasiado estrecha de experiencia” (NO I 62) y, por tanto, sistemas desdeñables; entre estos se incluyen los desarrolladas por Platón, Aristóteles, Pitágoras, entre otros autores.

La comprensión del concepto de forma en Bacon, dada la relevancia del conocimiento de formas para el conocimiento en general y su relación con la capacidad de modificar las propiedades de las cosas (las naturalezas simples), supone desafíos para el vínculo entre conocimiento y acción. No atender la noción de forma puede provocar que la propuesta metodológica del autor, la inducción baconiana, sea considerada como inadecuada para los fines de su filosofía por centrarse en el conocimiento de algo que a primeras aparece como oscuro y de todo menos concreto. Si no pueden relacionarse claramente el conocimiento de formas con la práctica, aquel debería dejar de ser interesante para el

<sup>1</sup>De aquí en adelante, *Novum Organum* será abreviado como NO. A la abreviatura le siguen el número del libro y el número del aforismo.

empirismo baconiano, pues el conocimiento en general está subordinado a la acción (Frondizi 2003). Perseguir un tipo de conocimiento que no informa significativamente nuestro actuar sería desde este punto de vista un desperdicio.

El presente escrito pretende examinar el concepto de forma en la obra de Bacon, determinar la contribución del conocimiento de formas a la empresa cognoscitiva general y determinar el valor de aquel frente a cuestiones como el conocimiento de las causas eficiente y material. Además, con pretensión de extender los frutos de la propuesta baconiana, se revisan brevemente posibles coincidencias con ideas contemporáneas y muy generales de la ciencia y su relación con la práctica humana.

## **2. El concepto de forma: una primera delimitación**

Francis Bacon fue un fuerte crítico de Aristóteles y la tradición posterior a él por el carácter sofisticado de sus desarrollos filosóficos y su creciente desapego de la experiencia para la reflexión sobre la naturaleza (Frondizi 2003). Sin embargo, en lo que concierne a la propuesta aristotélica de las cuatro causas, no se distanció por completo y decidió reformular tal distinción con el fin de aprovechar su extendido uso en la filosofía. Si bien incorporó con ciertas modificaciones los conceptos de causa formal, material y eficiente, se sirvió principalmente de este último tipo de causa.

Al examinar el concepto baconiano de forma, algunos intérpretes consideran que hay gran parecido con el concepto aristotélico de causa formal; sin embargo, Bacon parece tener en mente una noción más general que la de causa. En concreto, comprende la forma como (al menos) la condición esencial para la existencia de la naturaleza (Frondizi 2003, pág. 25), entendiendo aquí por naturaleza las cualidades o propiedades de un cuerpo dado. La aplicación de este concepto pareciera restringirse a las propiedades; quedan por el momento fuera de la investigación cuestiones como la forma general de la materia, de los cuerpos y de las sustancias. Serán consideradas dos maneras en que la forma es esencial para la existencia de la propiedad que le corresponde: esencial para la expresión natural de tal propiedad, o bien para la producción de esta en un cuerpo.

El conocimiento debe contribuir a que el hombre pueda modificar el mundo aumentando su dominio sobre la naturaleza mediante la comprensión de esta; después de todo “a la naturaleza no se la vence si no

es obedeciéndola” (NO I 3), y la ciencia debe proveernos de instrumentos para dominarla mediante un conocimiento adecuado de ella. Ahora bien, no todo elemento presente tradicionalmente en el conocimiento contribuye directamente a este objetivo. Por ejemplo, algunos axiomas (los de mayor generalidad) se desvinculan notablemente de los hechos particulares y de su observación, por lo que terminan siendo un poco abstractos (=no concretos) y de difícil aplicación práctica, aunque se muestren útiles para la expansión del conocimiento.

La ciencia es imperfecta si solo conoce las causas de las propiedades en determinados cuerpos y el poder es imperfecto si se puede producir un efecto solo en determinadas materias (es decir, solo en algunas dentro del conjunto de las susceptibles de tal efecto) (NO II 3). Estas deficiencias se presentan cuando solo se estudian las causas eficientes y materiales de los fenómenos. La única manera de lograr que la ciencia y el poder humano tengan más alcance y se apliquen a una colección diversa de cuerpos, fenómenos y materias, es mediante el conocimiento de las formas; de las condiciones esenciales presentes en todos los casos en que tal propiedad aparece (NO II 2-3). Una modificación de los cuerpos y sus propiedades requiere un conocimiento detallado del cuerpo en el que se trabaja (NO II 7).

En consideración de esta concepción de la forma, que es ambigua y poco detallada, las formas aparecen como algo medianamente abstracto<sup>2</sup>, es decir, como algo que por sí solo no indica cómo inducir o producir una naturaleza simple en un cuerpo, que no permite formular un curso de acción para la modificación del mundo. Además, si se considera lo abstracto como lo que no es concreto en cuanto las formas no son objetos materiales (Cameron 1964), también son en este sentido abstractas. Ahora bien, las formas (como sea que se conciban) siempre se vuelven inútiles si son comprendidas separadas de la materia, error que según Bacon parecen haber cometido los platónicos (NO I 65). Estos son rasgos de la forma que no parecen ser tan disputados. Sin embargo, aunque puedan reconocerse presentes en varias interpretaciones, la dificultad inicial sigue apareciendo: identificar qué tipo de objeto<sup>3</sup> son las formas y cómo se dan estos rasgos en un objeto es la

---

<sup>2</sup>Para Bacon, lo abstracto es lo que es inútil para guiar a los humanos a una modificación del mundo. Se considerarán además dos sentidos no baconianos del concepto: lo derivado indirectamente de los hechos particulares y lo que no es concreto.

<sup>3</sup>‘Objeto’ se utiliza acá en sentido lato, no como algo existente realmente.



cuestión central. A continuación se revisarán algunas interpretaciones disponibles que permiten responder esta pregunta.

### 3. Algunas interpretaciones del concepto de forma en Bacon

Recapitulando las ideas anteriormente revisadas: el conocimiento de formas aporta a la práctica humana en cuanto contribuye (parcialmente) a conocer los modos de producir ciertas naturalezas simples en diversos cuerpos. Esto es así en cuanto conocer la forma supone conocer las condiciones que han de cumplirse para la presentación de una naturaleza simple en un cuerpo. Ahora bien, no es claro cómo estas condiciones permitirían elaborar una regla o indicación para producir una naturaleza simple en una variedad de cuerpos, que es a lo que la ciencia aspira. El poder limitado a una pequeña clase de cuerpos muy semejantes entre sí es algo indeseable.

Así, la interpretación del concepto de forma se vuelve relevante pues permite vincular aquellas condiciones que vienen dadas en las formas conocidas con la acción humana. Se considerarán las siguientes interpretaciones:

1. Las formas son las leyes de la acción de la materia y las reglas para producir naturalezas simples en los cuerpos (Cameron 1964).
2. Las formas son las leyes que rigen las naturalezas simples (Frondizi 2003).
3. Las formas son principios de estructura corpuscular (Dickie 1923).
4. Las formas son estructuras constituidas por los elementos de la naturaleza (Klein 2020).
5. La forma es una naturaleza delimitada por otra que permite explicar una tercera naturaleza (Schwartz 2014).

En todas estas interpretaciones, la idea de que las naturalezas simples dependen en algún sentido de la forma se mantiene: o bien las formas son condición para la presentación de una naturaleza simple o son requeridas como conocimiento para la producción de la propiedad en los cuerpos. En este sentido, son “la condición esencial para la existencia de la naturaleza” (Frondizi 2003, pág. 25). La diferencia relevante entre estas interpretaciones está en lo directo que es el vínculo entre el conocimiento de formas y la producción de naturalezas simples.

En el caso de las lecturas de Dickie (1923) y de Klein (2020), el problema de la relación entre el conocimiento de formas y la práctica es más evidente que en las opciones restantes. Dickie (1923) comprende las formas como principios de estructura corpuscular, como aquellos enunciados generales sobre cómo los corpúsculos se organizan y generan estructuras regularmente. Esta lectura cobra sentido considerando que Bacon sostiene que toda alteración de un cuerpo supone un movimiento de sus partículas constitutivas (NO I 51, nota 50). La producción de una naturaleza simple, por tanto, supone un movimiento de tales partículas. Ahora bien, el que podamos conocer los principios que suele seguir la estructura u organización corpuscular en cuerpos con ciertas naturalezas simples no necesariamente permite que sepamos cómo modificar tal estructura u organización. Por tanto, conocer tales principios dados en ciertos cuerpos con determinadas naturalezas simples no nos garantiza el poder modificar tales estructuras para producir tales naturalezas. El problema inicial de encontrar la manera de inducir (=producir) una naturaleza simple en un cuerpo se vuelve también el problema de encontrar la manera de cambiar su estructura corpuscular.

La interpretación de Dickie (1923) presenta otros problemas. Considérense lo mencionado por Frondizi (2003) sobre el esquematismo latente (Libro I 51, nota 49): este es la manera en que están ordenadas las pequeñas partículas que constituyen los objetos materiales y es latente en cuanto es inaccesible a la observación. Surge entonces la pregunta por cómo sería posible hablar de principios de estructura corpuscular si ni siquiera podemos observar aquella estructura. Bacon tenía la expectativa de que la ciencia pudiese conocer en algún momento los esquematismos y procesos latentes y así eventualmente encontrar la manera de modificarlos; solo en tal escenario la cuestión sobre cómo inducir una naturaleza simple podría ser respondida mediante principios de estructura corpuscular. Ahora bien, considerando que este tipo de observación no era posible en la época de Bacon, suponer que esta interpretación es adecuada sería aceptar que el mismo Bacon propuso un concepto clave en la investigación de la naturaleza que impedía la misma investigación al carecer del componente de observabilidad.

En el caso de Klein (2020), las formas son consideradas como las estructuras que los elementos de la naturaleza constituyen, es decir, las estructuras formadas por las partículas elementales (*simplicitatem naturae* en NO I 57). El problema que se presenta a partir de esta lectura es similar al de la interpretación de Dickie (1923): es difícil conceder que Bacon, considerando la ciencia de su época, haya aceptado

abiertamente que uno de los principales elementos del conocimiento fuera inobservable en su tiempo. Ambas dependen de una física no disponible para Bacon, cuyos instrumentos permitirían la observación del esquematismo y proceso latente y su posterior generalización en principios o estructuras. Ahora bien, esta observación no es incuestionable, pues funciona solo suponiendo que en general la obra de Bacon es coherente y está fuertemente atravesada por su empirismo, de modo que difícilmente incorpora elementos centrales que no estén disponibles a los sentidos, ni siquiera indirectamente mediante sus efectos.

Si Bacon sostiene que sus formas no son abstractas, sino que son útiles para la acción humana, entonces estas deben referir, sea como principios o estructuras, a cuestiones modificables. Ante la imposibilidad de observar el esquematismo y el proceso latente, estos elementos no son modificables de manera controlada sin la mediación de las naturalezas simples como indicadores. El modo de producir una naturaleza simple desde la perspectiva baconiana siempre debe ser descrito respecto de algo observable y modificable como las naturalezas simples. Ahora bien, otro asunto es el problema que aparece al intentar vincular indicadores con estructuras inaccesibles e indistinguibles para los humanos.

En cambio, las interpretaciones de Cameron (1964), Frondizi (2003) y Schwartz (2014) presentan dos virtudes frente a las de Dickie y Klein. En primer lugar, son interpretaciones aparentemente más fieles y compatibles con el texto en cuanto consideran: (i) la incapacidad de la ciencia de la época de acceder mediante la observación a las partículas elementales, su esquematismo y su proceso latente; (ii) el carácter medianamente abstracto (=no concreto, general) de las formas y su pertenencia a la metafísica como objeto de investigación; y (iii) la insuficiencia del conocimiento de formas para la producción de naturalezas simples (NO II VII). Además, permite que la relación entre conocimiento de formas y práctica sea más directa, lo que se revisará a continuación.

La interpretación de Frondizi (2003) supone comprender las formas como aquellas leyes que rigen sobre las naturalezas simples, como aquellos enunciados generales descriptivos de la relación entre diversas naturalezas. Estas leyes permiten relacionar la presencia de una naturaleza simple (como el calor) con otras naturalezas (como el movimiento). De este modo, en cuanto enunciados, permiten vincular la presencia y producción de una naturaleza como condición para la presencia y producción de otra. Este vínculo no es del tipo causa-efecto,

sino de condición y posibilidad derivada del cumplimiento de tal condición (Fronidzi 2003).

Cameron (1964), que destaca el elemento pragmatista en la obra de Bacon, entiende las formas como las leyes de la acción de la materia y las reglas para producir naturalezas simples. Es decir, como enunciados generales descriptivos de cómo la materia suele comportarse o bien que muestran el origen de una naturaleza simple en términos de otra. A pesar de que Bacon considera que toda alteración en un cuerpo supone movimiento en sus partículas, esta noción de la forma no parece poner énfasis en la estructura corpuscular, sino en los cambios que pueden observarse en la materia y los cuerpos<sup>4</sup>. La ley de la acción de la materia, por tanto, vincularía efectos observables de cambios en la materia que pueden ser identificados como naturalezas simples. El vínculo entre las naturalezas relacionadas mediante esta ley muchas veces es del tipo género-especie, como Bacon indica a continuación:

[...] he dicho acerca del movimiento (a saber que es a modo de género respecto a su especie el calor) entendiéndose no que el calor engendre movimiento o que el movimiento engendre calor (aunque esto es cierto en algunos casos) sino que el calor mismo (*ipsissimum*) o el *quid* mismo (*quid ipsum*) del calor es movimiento y nada más [...]. (NO II 20)

Considerando las formas como reglas para producir naturalezas simples, estas indicarían qué naturaleza ha de estar presente y debe ser inducida en el cuerpo para lograr la aparición o posibilitar la producción de una segunda naturaleza. Si se sabe cómo producir la naturaleza requerida para inducir la naturaleza que es objeto de estudio, entonces se conoce la forma de la naturaleza estudiada, pues se conoce la regla que es eficiente para la inducción de la naturaleza. En este sentido, todas las formas son útiles pues permiten la producción de naturalezas.

Finalmente, considérese la interpretación de Schwartz (2014). Esta lectura propone que la forma es una naturaleza delimitada por otra que permite explicar una tercera naturaleza. Es decir, si se quiere estudiar la naturaleza simple del calor, la forma de esta sería la naturaleza del movimiento delimitada por la naturaleza 'expansiva' de modo que la

<sup>4</sup>Esta interpretación considera observable incluso el proceso latente en cuanto escapa parcialmente a los sentidos, pues sus efectos al nivel de las naturalezas simples sí son visibles.

forma del calor sería el movimiento expansivo. Es decir, entre los tipos de movimiento, el calor en cuanto especie del género del movimiento es precisamente un movimiento de carácter expansivo. Esto indica que al producir la naturaleza del movimiento en un cuerpo con el fin de producir la naturaleza del calor, el resultado de la inducción (=producción) del movimiento debe ser un movimiento expansivo y no de otro tipo. Las formas, sin embargo, no son suficientes por sí solas y han de ser consideradas junto a la causa eficiente y material, ofreciendo así guía práctica y expandiendo el dominio de objetos sobre los cuales se puede producir una naturaleza simple. Tales causas, según Schwartz (2014), deben ser estudiadas por la física para facilitar la producción de naturalezas simples sobre cuerpos realmente susceptibles.

Frente a las interpretaciones de Dickie (1923) y de Klein (2020), que refieren a las partículas y su organización, surge la pregunta por cómo la inducción baconiana, que inicia siempre desde enunciados a partir de la observación, podría partir de un enunciado sobre algo inobservable, en este caso, las partículas y su esquematismo. El único modo en que la inducción podría derivar en principios de estructura corpuscular es mediante la observación del esquematismo. Los principios de estructura corpuscular o las estructuras constituidas por las partículas serían justamente reflexiones sobre observaciones de tales partículas; sin embargo, estas no eran observables para la ciencia de la época de Bacon. Si las formas de Bacon son alcanzadas mediante inducción baconiana, que depende de la observación en cuanto es su material de trabajo, y estas son principios de estructura corpuscular o la estructura constituida por las partículas, las formas serían principios generalizados a partir de observaciones de cuestiones inobservables, lo que es evidentemente problemático.

En cambio, las otras tres interpretaciones presentadas parecen ser más compatibles con la idea de inducción baconiana al considerar las leyes como generalizaciones de cuestiones observables (i.e., naturalezas simples) o como generalizaciones de los intentos de inducir una naturaleza simple y los resultados de estos. Esto permitiría que las formas así entendidas puedan ser alcanzadas mediante un proceso inductivo que sí parte de la observación.

Además de su compatibilidad con el *Novum Organum*, estas lecturas aparecen como convenientes en lo que respecta al concepto de forma, al mantener el vínculo de la inducción de naturalezas con la experiencia. Si las formas son leyes que rigen las naturalezas simples o generalizaciones descriptivas que relacionan una naturaleza con otra,

estas funcionan como reglas para inducir una naturaleza simple, siempre complementadas por la información que la física pueda brindar respecto de la causa eficiente (proceso latente) y material (esquematismo latente). La interpretación de las formas entendidas como reglas para producir propiedades se ve respaldada por afirmaciones de Bacon como la siguiente:

Ciencia y poder humanos coinciden en una misma cosa, puesto que la ignorancia de la causa defrauda el efecto. A la naturaleza no se la vence si no es obediéndola, y lo que en la observación es como causa es como regla en la práctica. (NO I 3)

Incluso aceptando estas interpretaciones, es necesario considerar que Bacon nunca concede a las formas el poder de permitir la inducción universal de naturalezas simples. Es decir, no por descubrir la forma se conoce con certeza la manera de inducir la naturaleza simple correspondiente sobre cualquier cuerpo o cualquier materia. Tampoco las formas por sí solas pueden entregar una regla para la inducción de naturalezas simples en los cuerpos, sino que el conjunto de las formas, el esquematismo y el proceso latentes permitirían un mejor dominio del mundo, aunque no sea perfecto (Cameron 1964). Considerar solo las causas material y eficiente sería conformarse con modificar cuerpos materialmente muy semejantes y, por tanto, conformarse con reducir el poder humano notablemente; por ello, el conocimiento ha de aspirar a conocer las formas para expandir su dominio de la naturaleza.

#### **4. Reconsiderando el vínculo entre ciencia y práctica en la filosofía baconiana**

Para Bacon, la ciencia natural se divide en dos disciplinas: la física y la metafísica. La física se encarga de investigar las causas particulares “la causa eficiente y de la materia, del proceso latente y del esquematismo latente” (NO II 9); no se encarga de las “leyes fundamentales y eternas” (NO II 9). La metafísica, en cambio, reflexiona en torno a las causas generales y constantes (Klein 2020), en torno a las formas eternas e inmutables (NO II 9). En cuanto ambas son parte de la ciencia natural, la física y la metafísica son hijas de la madre común de todas las ciencias: la filosofía natural (NO I 79).

La relación entre estas dos áreas de la ciencia se da de manera tal que, a partir de las observaciones recopiladas en la historia natural, la

física investiga las causas de los fenómenos descritos, llegando a relaciones invariantes hasta finalmente alcanzar correlaciones más inclusivas y finalmente desembocando en las formas, que son asunto de la metafísica (Klein 2020). La ciencia, según lo planteado por Bacon, se ocuparía mediante este procedimiento de “dotar a la vida humana de nuevos inventos y recursos” (NO I 81).

Ahora bien, la relación entre ciencia natural y práctica humana no es tan comprensible si solo se considera el conocimiento de las formas requerido por la ciencia natural como base para la derivación de reglas para la acción e inducción de naturalezas simples. Dada la generalidad de las formas, cursos específicos de acción orientados a determinados efectos en los cuerpos difícilmente pueden ser derivados de estas leyes. Para planificar la acción humana, parece adecuado incluir también lo que Bacon llama axiomas medios, que son “los axiomas verdaderos, sólidos y vivos, en los que se asientan las cosas e intereses humanos” (NO I 104) y que, según Frondizi (2003), “unen los datos inmediatos de las cosas individuales con sus leyes abstractas universales” (NO I 13, nota 17), es decir, los unen con las formas.

La incorporación de elementos como los axiomas medios permite que las formas sean un componente de la filosofía baconiana compatible con el empirismo del autor, a la vez que permiten dar cuenta de un modo controlado, no precipitado, de expansión del conocimiento a partir de la observación. Bacon parece haber reconocido abiertamente la necesidad de generalización en la ciencia natural, la posibilidad de alcanzar legítimamente aquellos enunciados generales sin abandonar por completo lo dado en la experiencia y la aplicabilidad no irrestricta de estas generalizaciones (el conocimiento de la forma no garantiza la posibilidad de inducir universalmente la naturaleza simple).

Estos aspectos de su propuesta le permiten rescatar partes de los proyectos de filosofía natural disponibles en su tiempo, no abandonar el ideal de una ciencia que tiene conocimiento/dominio sobre clases amplias de objetos (no solo sobre algunos objetos particulares) y conectar con algunas de las pretensiones de la ciencia contemporánea, o al menos con las concepciones de la ciencia que reconocen la necesidad de “ir más allá de la experiencia” en un sentido amplio, empleando el intelecto de manera más “segura”. La filosofía baconiana no pretende para la ciencia un método que le entregue certezas absolutas, sino solo más certeza o bien mayor utilidad para la vida humana, mayor capacidad para mejorar el estado del hombre. Los frutos del conocimiento

son para Bacon los indicadores de la verdad de una teoría (NO I 73), y esto también es una expresión clara de la relación ciencia-práctica.

En conexión con esta idea de la utilidad de la ciencia, Bacon sostiene que la ciencia contribuye y debe contribuir con inventos y recursos para la vida humana (NO I 81) para aumentar el dominio del hombre sobre la naturaleza y ponerla a su servicio. El carácter de estos inventos y recursos no es delimitado claramente por el autor, de modo que será considerado de manera laxa, incluyendo herramientas ‘conceptuales’ y ‘prácticas’. Es de particular interés el caso de las herramientas prácticas, pues cabe preguntarse si entre los inventos prácticos de la ciencia natural pueden incluirse los artefactos, es decir, aquellos “objetos que pueden controlarse y ser usados para fines específicos” (Romero 2018, pág. 66), usualmente producidos por la tecnología para aumentar o mejorar nuestro control sobre el mundo. Lo mismo puede cuestionarse respecto de la planificación de la actividad humana para el control de ciertos procesos, sobre todo si las reglas para inducir naturalezas se consideran como este tipo de planificaciones. Considerando la interpretación de las formas como reglas para la producción de naturalezas, solo puede afirmarse que al menos parte de estos inventos corresponde a indicaciones para producir que una propiedad se presente en un cuerpo. Averiguar si en la obra de Bacon se presenta, aunque sea seminalmente, un concepto semejante al de tecnología que permita el manejo y la especificación práctica de las normas, y delimitar el carácter de estos inventos/recursos podría ayudar a precisar aún más la relación entre ciencia y práctica, sobre todo si se concede que la tecnología, como disciplina actualmente informada por el conocimiento científico, ha permitido mediar más eficientemente entre estos dos ámbitos (aunque tal mediación no venga libre de problemas).

Una revisión más sistemática y comprensiva de la obra de Bacon más allá del *Novum Organum* y en consideración de la totalidad de su proyecto permitiría determinar si se presentan versiones más simples de conceptos como el de tecnología y a esclarecer más la conexión entre ciencia y práctica dentro de lo que la obra de Bacon permite. Estas tareas no son solo de interés para la historia de la filosofía, sino que pueden reportar resultados para la reflexión en torno a la idea de que la práctica humana, aunque sea en un sentido general, puede verse informada por la ciencia, llevando sus conocimientos generales a acciones particulares; sobre todo considerando el contexto filosóficamente fragmentado en que Bacon sostuvo esta innovadora opinión.



## Referencias

- Bacon, F. (2003). *Novum Organum*. (C. F. Almorí, Trad.). Losada.
- Cameron, E. W. (1964). Francis Bacon and the Pragmatic Theory of Forms. *Philosophical Forum*, pp. 592-610. <https://api.core.ac.uk/oai/oai:yorkspace.library.yorku.ca:10315/37086>
- Dickie, W. M. (1923). "FORM" AND "SIMPLE NATURE" IN BACON'S PHILOSOPHY. *The Monist*, 33(3) pp. 428-437. <http://www.jstor.org/stable/27900958>
- Fronzizi, R. (2003). Significado y contenido del *Novum Organum*. En F. Bacon, *Novum Organum* (C. F. Almorí, Trad., págs. 9-35). Losada.
- Klein, J. (2020). Francis Bacon. Obtenido de Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/archives/fall2020/entries/francis-bacon>
- Romero, G. E. (2018). *Scientific Philosophy*. Cham: Springer Verlag. <https://doi.org/10.1007/978-3-319-97631-0>
- Schwartz, D. (2014). Baconian Foundationalism and the Problem of Certainty. <https://escholarship.org/uc/item/7r33h8t7>



# Conjuntos de creencias y coherentismo: posibles manipulaciones del conocimiento científico

Miguel Naredo Rojas<sup>a</sup>

## Resumen

Ante un mundo tecnológico en el cual lo que prima es la ciencia, se vuelve menester hacer un análisis crítico de la misma. Con esta premisa, tiene origen el escrito que hoy presentamos. Nuestro objetivo será buscar mecanismos y esquemas que nos permitan evaluar y comprender el funcionamiento de la ciencia, y, sobre todo, cómo se toman decisiones dentro de ella. Entender cómo se establece el conocimiento científico y qué lo justifica es un eje central dentro de esta explicación. Con ello, nuestra intención será proponer un esquema para evaluar y detectar ciertas situaciones irregulares dentro de la ciencia; las cuales van en contra de la visión que se ha popularizado de la misma, e incluso de la perspectiva que la comunidad científica quiere dar. Nuestro cometido hoy es desmitificar la ciencia, exponer su estructura y desarrollo, con la intención de permitir una visión más cercana a la práctica científica a mano de toda persona que se lo proponga. Para hacernos con esa herramienta de análisis anteriormente nombrada partiremos del conocimiento científico y su justificación; para después centrarnos en un modelo en específico: el coherentismo. A partir del mismo, exploraremos cómo se toman ciertas decisiones que escapan de lo que consideraríamos “científico” y dan lugar a ciertas situaciones que parecen anómalas dentro de la disciplina; y, posteriormente, dan lugar a manipulaciones del conocimiento.

**Palabras Clave:** Conocimiento, Coherentismo, Justificación, Paradigma, Ciencia normal.

---

<sup>a</sup>Universidad de Oviedo, Oviedo, España.

Contacto: UO288715@uniovi.es

## Abstract

In a technological world where science prevails, it becomes necessary to conduct a critical analysis of it. With this premise, the writing we present today originates. Our goal is to seek mechanisms and frameworks that allow us to evaluate and understand how science operates, and above all, how decisions are made within it. Understanding how scientific knowledge is established and what justifies it is a central axis of this explanation. With this in mind, our intention is to propose a framework to assess and detect certain irregular situations within science, which go against the popularized view of it, and even the perspective the scientific community aims to project. Our mission today is to demystify science, expose its structure and development, with the intention of offering a vision of scientific practice that is more accessible to anyone who seeks it. To equip ourselves with the aforementioned tool of analysis, we will begin with scientific knowledge and its justification, and then focus on a specific model: coherentism. From this, we will explore how certain decisions are made that stray from what we would consider “scientific,” and lead to situations that appear anomalous within the discipline, and subsequently give rise to manipulations of knowledge.

**Keywords:** Knowledge, Coherentism, Justification, Paradigm, Normal Science.

## 1. La Ciencia como Paradigma del conocimiento

En el mundo contemporáneo hemos llegado a lo que para algunos es el culmen del desarrollo humano, la punta de lanza del progreso (Comte 1844), aquel faro que por fin nos iluminará el oscuro mundo que nos rodea; en pocas palabras, nos encontramos en la época científica donde la mística, la mitología y la teología han sido sustituidas por las teorías científicas, los avances tecnológicos y la lógica. La razón ha devorado cualquier atisbo de fantasía y con ello ha establecido un nuevo paradigma epistemológico, una nueva forma de comprender todo aquello que nos rodea. Ahora, todo puede ser reducido a un conjunto de proposiciones que cumplan ciertas directrices establecidas por el propio paradigma científico: que sea sólido, congruente, lógicamente válido, pertinente y empíricamente comprobable. Algunos se atreverán a decir que nos encontramos ante una nueva religión, otro nuevo dogma que imparte su forma de comprender el mundo con la ligera diferencia de que no recurre a entes sobrenaturales para convencer a sus adeptos, sino que se aprovecha de la ignorancia de las personas para hacerlas creer en complejas teorías completamente incomprensibles para cualquier ciudadano de a pie, que no le queda otra opción más que llevar a cabo un acto de fe al creer en ellas. No es nuestra intención juzgar la

validez de las teorías científicas, tampoco lo es adoptar una postura tan radical como la anteriormente nombrada. Poco provechosa sería nuestra labor si redujéramos a la ciencia y su conocimiento a otra secta más con un nuevo discurso. En cambio, nos interesa la ciencia como nuevo paradigma, como el actual medio de comprensión y reconocimiento del mundo, aquella fuente de conocimiento fidedigno en la cual confiamos.

Aunque no vamos a declarar la guerra a la ciencia, tampoco pensamos rendirle tributo ni ser su más fiel adepto; nuestro objetivo es hacer visibles ciertos trastornos que pueden surgir en la elaboración y construcción de las distintas teorías y conjuntos de creencias científicas en cuanto durante su desarrollo pueden causar manipulación y prejuicios o daños epistémicos (Fricker 2007) que nublen nuestra concepción del mundo a partir de elecciones arbitrarias al establecer el conocimiento científico. De tal manera que nos preguntaremos: ¿la ciencia siempre tiene un criterio objetivo? Y si lo tiene, ¿toda decisión llevada a cabo dentro de la ciencia no se ve influenciada por algún otro factor que no sea científico, epistémico o lógico? Sumado a ello, también buscamos romper con ciertos mitos dentro de la ciencia que son muy comunes y que contribuyen a una percepción errónea del conocimiento y la comunidad científica en sí. Volver explícitas las motivaciones de la investigación científica será clave para comprender las decisiones que se toman dentro de la comunidad. A su vez, conocer qué individuos o conjunto de individuos tiene poder dentro de la ciencia al momento de tomar decisiones será uno de nuestros objetivos. Veremos qué tan cierto es el mito del “científico encerrado en su despacho, únicamente dedicado a su investigación” (Baudelaire 1857) como si de un monje se tratase, completamente alejado de implicaciones sociales de cualquier tipo y únicamente centrado en la ciencia. Como si la ciencia fuese cosa solo de científicos.

Si nos planteamos concebir la ciencia como el nuevo paradigma —nuestra herramienta más fiable y método para describir el conglomerado de nuestra existencia y lo que conocemos— debemos profundizar en conceptos de interés en nuestra labor. Haremos un breve recorrido por la noción de conocimiento y cómo se establece. A su vez, exploraremos qué es un conocimiento válido y cómo se justifica. Con ello, una vez tengamos una idea clara de cómo se estructura el conocimiento, nos adentraremos en cómo se estructura el conocimiento y sus manipulaciones. No pretendemos atacar a una ciencia dogmática y sectaria, sino contra la posibilidad de que se vuelva aquello que acabamos de nombrar.

## 2. Conocimiento y Conjunto de Creencias

Previo a iniciar con la exposición de nuestros argumentos, cabe destacar lo siguiente: trabajaremos principalmente con una noción de conocimiento relacionada con la ciencia. También veremos las distintas definiciones de nuestro interés a lo largo de la empresa científica. Con ello, buscamos trazar un camino a seguir dentro de nuestra explicación para a su vez abordar todas aquellas definiciones pertinentes para nuestro análisis.

En general, la ciencia busca establecer conocimientos comprobables. Dentro de su campo pretende explicar, describir y predecir diversos fenómenos del mundo (sociales, naturales, etc.) (Hempel 1996). Para ello cuenta con un método propio y específico mediante el cual se confirman y establecen los conocimientos (Hempel 1996, p. 247). Pero, ¿cómo sabemos que un conocimiento está comprobado? Usualmente creemos que verificar un conocimiento es poder demostrarlo en la realidad, aunque seguimos sin saber bien cómo se demuestra ni qué significa el demostrarlo en sí. Para despejar las dudas hemos de definir lo que es el conocimiento. Con dicho fin acudimos a Platón (*Teeteto* 201c-210b) donde el conocimiento se define como creencia verdadera justificada.

El conocimiento es una creencia en tanto es un estado mental previo al conocimiento y necesario para él —no podemos conocer nada de lo que no creamos—; es una creencia verdadera en cuanto debe corresponderse a la realidad —debe existir una correspondencia entre mi creencia y el mundo— y por último se justifica ya que existe la posibilidad de tener una creencia verdadera por diversas razones, incluso por suerte, de tal manera que salvo que podamos justificar nuestra creencia verdadera, no conocemos.

Ahora sabiendo qué es el conocimiento podemos volver a nuestra pregunta: ¿Cómo sabemos que un conocimiento está verificado? Tras lo dicho anteriormente, los más sagaces dirán que por la justificación, para poder saber si una creencia verdadera es un conocimiento habrá que comprobar si su justificación es suficiente para considerarlo conocimiento. Esta noción del término perduró durante mucho tiempo; no fue hasta la publicación de “¿Is Justified True Belief Knowledge?” de Edmund Gettier (1963) en donde se nos presentan diversas situaciones en las que la definición de conocimiento dada es insuficiente. Si bien justificamos las creencias verdaderas para escapar de la incertidumbre y de tener creencias por mera suerte, las propias justificaciones también son susceptibles de ser por mero azar o suerte. Como bien sabemos, al-

go que llegamos a saber por mera suerte realmente no lo conocemos, de tal manera que nuestra pregunta sigue abierta pero ahora sabemos dónde podemos hallar la solución.

Como hemos visto, la creencia verdadera justificada no es un criterio adecuado para describir el conocimiento, pues la fortuna sigue teniendo un papel de relevancia en la misma. Si nos queremos deshacer de esta, necesitaremos otro criterio más aparte de los que ya tenemos. Si podemos tener una creencia verdadera por suerte, al igual que podemos justificarla por puro azar, tendremos que buscar una relación directa entre la verdad de la creencia y su justificación. Esta relación nos permite unir directamente ambos criterios y evita la posibilidad de llegar a ellos por azar; dado que están directamente relacionados, la justificación de la creencia será parte de su verdad y viceversa. Existe una gran variedad de posturas para este nuevo criterio y cómo debería ser, pero no es de nuestro interés pasar por todas ellas para esta exposición; en cambio nos centraremos en una que ha tenido gran relevancia dentro del conocimiento científico, siendo una de las más apoyadas como cuestionada a su vez: el coherentismo (Darcy 2007 , pp. 134-160).

En principio se debe comprender que el coherentismo es una postura internalista (Darcy 2007 , pp. 64-66, 153-164); esto quiere decir que el criterio que buscamos para fortalecer nuestra definición de conocimiento se encuentra dentro del individuo conocedor y sus estados mentales. Por consiguiente, el coherentismo trata sobre las creencias y su justificación a nivel mental. El coherentismo, a grandes rasgos, sugiere lo siguiente: las creencias verdaderas están justificadas en cuanto son parte de una relación holística —como un todo— entre otras creencias dentro de un conjunto (Darcy 2007 , pp. 132-143). En otras palabras, una creencia está justificada cuando es coherente con el conjunto del que forma parte. El coherentismo busca que exista una concordancia entre las creencias y que sean coherentes entre sí. También hay que tener en cuenta que los conjuntos deben ser lógicamente consistentes. Además, es necesario que el coherentismo defina en qué consiste la coherencia dentro de los conjuntos de creencias, es decir, hasta qué nivel algo es coherente. El coherentismo busca que una creencia tenga concordancia y sea coherente con un conjunto. Sin embargo, cuantas más creencias apoyen a la creencia verdadera a justificar, más fuerte será esta justificación. Si queremos establecer un límite de creencias conectadas que pueden existir dentro de los conjuntos, este se vuelve difuso y vago. A pesar de ello, para la justificación de las creencias esto no es de gran importancia.

Esta postura es algo ambigua pues traza pocos límites de lo que podría ser un conjunto de creencias válido; si tomamos el coherentismo como la vía legítima de justificación, no escapamos de las inconsistencias pues cualquier conjunto de ideas bien formado puede ser coherente (incluso un cuento). Sumado a ello, las personas en su vida cotidiana abordan el conocimiento en función del que ya tienen. Estas dos situaciones dan como resultado que el coherentismo recaiga fácilmente en relativismo (Young 2007). Como podemos observar, el coherentismo tiene que hacer frente a diversos problemas, entre ellos podemos encontrar el de la pluralidad (Hintikka 2007). No tenemos una definición de coherencia que evite que varios conjuntos, siendo igualmente coherentes y válidos, representen diversas verdades sobre un mismo fenómeno. Si suponemos que solo puede existir un conjunto válido en concordancia con el principio de no contradicción (Aristoteles *Metafísica*: IV 3 1005b19-20), el coherentismo debe poder establecer su propio criterio de verdad para poder limitarse a ciertos conjuntos verdaderos y evitar las contradicciones. Esto lleva a que el coherentismo caiga en un fundacionalismo (Young 2003) al establecer su propio criterio de verdad sobre los conjuntos de forma arbitraria. Se suele atribuir a la lógica este criterio de coherencia y verdad en cuanto conjunto, pero al poder encontrarse con situaciones como las anteriormente mencionadas, observamos que se vuelve muy compleja la justificación de creencias desde el coherentismo.

Dentro de lo mencionado anteriormente se encuentra una de las principales tesis de este escrito; la que vale la pena destacar ya que será nuestra principal herramienta dentro de la argumentación. Esta es la siguiente: si tomamos el criterio de verdad fundacionalista dentro del coherentismo recaemos en universalismos arbitrarios que nos den como resultado un marco teórico con valores de verdad totalmente inconsistentes. Como resultado no existe una justificación real para elegir un conjunto de creencias u otros, de tal manera que ante dos justificaciones válidas de un mismo fenómeno no existe un criterio de elección que me permita decantarme por una u otra. Finalmente, pese a que justifique mi creencia, realmente no tengo una justificación para mi elección.

Si esto sucede, podemos encontrarnos con casos en los que el conocimiento científico sea completamente arbitrario; incluso siendo elegidas a conveniencia las justificaciones que sean beneficiosas para ciertos colectivos. Dadas estas situaciones es posible que nos veamos envueltos en perjuicios y daños epistémicos dentro de un campo que busca todo lo contrario. Además, pueden darse casos en los que las justificaciones



se elijan convenientemente para algún individuo o individuos, lo que resultaría en una descarada manipulación del conocimiento. La situación sería tal que se estaría justificando la elección de una justificación específica por el beneficio que puede proporcionar a quien la elija. Cuando se producen situaciones como las anteriormente mencionadas, en las cuales el criterio de elección de justificación es arbitrario, se permite la producción de conocimiento sesgado. Así, solo se divulgará el conocimiento que sea de interés para los grupos de investigación y divulgación científica según sea el caso. Tras esta manipulación del conocimiento, si se prolonga en el tiempo, se darán grandes conjuntos de conocimiento completamente manipulados, basándose unos en otros y dejando de lado todo lo que no beneficie a los grupos que promovieron este conocimiento. Posteriormente revisaremos qué impulsos tiene la ciencia extrínsecamente a su funcionamiento y qué relaciones establece con el mundo para comprender de qué manera pueden ser provocados estos trastornos.

Para continuar con nuestra exposición sobre cómo puede surgir la manipulación en el conocimiento científico en cuanto se pueden elegir conjuntos de creencias de forma arbitraria, observemos el siguiente ejemplo. Podemos preguntarnos lo siguiente: si bien ya nadie se atreve a decir que el tabaco no es dañino, es más difícil establecer hasta dónde llegan sus daños. Entonces, ¿qué tan dañino es el tabaco? Buscando establecer un conjunto de creencias que delimite qué tan perjudicial es para la salud, consideremos la siguiente proposición:

$p$  = El tabaco es dañino.

Tenemos dos conjuntos de creencias: un conjunto A donde el tabaco no es muy dañino, causando enfermedades leves y daños de poca importancia; y un conjunto B donde el tabaco es perjudicial para la salud, provocando cáncer y diversos problemas de gravedad para los consumidores. Dado este caso, puedo justificar mi creencia “ $p$ ” dentro de los 2 conjuntos, ya que ambos son epistémicamente válidos, pero el criterio de elección cambiará en cuanto el individuo elija la justificación. Por ejemplo, una tabacalera, por sus intereses comerciales, optará por el conjunto A para justificar su creencia “ $p$ ”; en cambio, un ex-consumidor de tabaco, por sus experiencias personales, considerará más adecuada justificar la creencia “ $p$ ” con el conjunto de creencias B, dando cada uno lugar a su propia manipulación del conocimiento.

Tras exponer el ejemplo, lo que interesa resaltar del mismo es el esquema al que da lugar. Supongamos que tenemos cualquier creencia

X la cual puede ser justificada a través de dos conjuntos distintos y contrarios entre sí —son mutuamente excluyentes—. Al elegir un conjunto como justificación el otro es descartado instantáneamente. Es en este instante cuando se elige alguno de los conjuntos donde surge la manipulación.

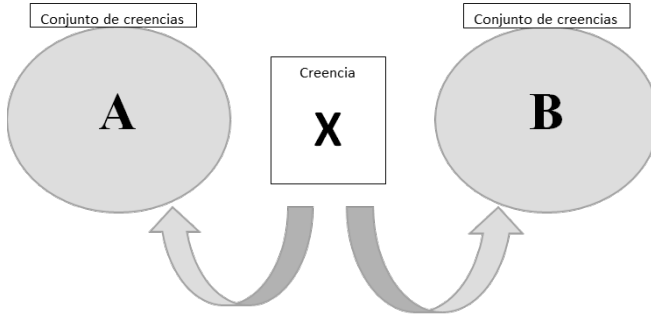


Figura 3: Esquema de justificación de creencias en conjuntos igualmente válidos y contrarios. Momento exacto donde surge la manipulación. (elaboración propia).

Esto nos deja de manifiesto lo siguiente:

- A falta de un criterio objetivo en el coherentismo para elegir conjuntos válidos la elección es meramente arbitraria.
- Si la elección es arbitraria, dentro de la ciencia existen diversos impulsos que afectan esta decisión más allá de la propia disciplina científica.
- Al ser elegidos los conjuntos de creencias de forma instrumental recaemos en un dogmatismo arbitrario y en una manipulación de conocimiento “válido”.

### 3. Los Paradigmas de la Ciencia y algunos mitos

Hemos planteado situaciones donde la manipulación da lugar a conocimientos sesgados, arbitrarios y manipulados; pero nos limitamos únicamente al conocimiento que es producto de la ciencia: conjuntos de creencias, teorías e hipótesis científicas. Pese a ello, aún no agotamos la totalidad de las formas que nuestro modelo puede explicar: otro conocimiento o mejor dicho, otra narrativa que se puede ver

oscurecida por el funcionamiento de la ciencia en su historia y desarrollo. Contar una historia de la ciencia es una labor ardua y de gran dificultad, principalmente por lo particular de su discurso, pues se concibe a la ciencia como una diagonal en constante crecimiento dentro de una gráfica, siempre avanzando y constantemente progresando (Comte 1980). Resulta curioso que dentro de la ciencia se cuenten fantasías como esta. Observaremos cómo la ciencia se va desarrollando a lo largo del tiempo, cómo se ha estructurado y qué clasifica los cambios a los que llamamos ‘progreso’. También cuestionaremos hacia dónde realmente apunta esta curiosa idea, al igual que rastrearemos qué motivaciones existen tanto dentro como fuera de la ciencia en el desarrollo científico. Por último, cuestionaremos cómo se narra la historia de la ciencia y qué tanto corresponde al verdadero desarrollo de esta.

Partiendo del esquema establecido anteriormente, podemos encontrar ciertas relaciones a lo largo de la historia del desarrollo científico. Primero habrá que establecer a lo que nos referimos con paradigma científico: con ello nos remitimos a los conjuntos de teorías y conocimientos, herramientas y dinámicas establecidas para el desarrollo de la ciencia normal (Kuhn 2011 , pp. 33-34); y con estos hacer frente a las diversas discrepancias y problemas que surjan entre el mundo y el conocimiento científico. La ciencia normal es un término acuñado por Kuhn (1962) el cual define de la siguiente manera ‘ciencia normal’: “investigación basada firmemente en una o más realizaciones científicas pasadas, realizaciones que alguna comunidad científica en particular reconoce durante cierto tiempo como fundamento para su práctica posterior” (Kuhn 2011 , pp. 224-246). Para comprenderlo un poco mejor observemos algunos ejemplos de paradigmas de gran relevancia histórica: el modelo geocéntrico y el modelo heliocéntrico (Fundación Gustavo Bueno 2017). Son dos claros ejemplos de lo que son paradigmas científicos. Ambos tienen sus propios conjuntos de creencias y conocimientos que condicionan el desarrollo de la ciencia normal, dando lugar a percepciones completamente distintas del mundo y siendo contrarios entre sí. Una puntualización de vital importancia es que dos paradigmas contrarios nunca conviven en un mismo momento; con ello se sobreentiende que al existir dos paradigmas para un mismo conjunto de fenómenos se debe decidir entre uno de los dos. Por consiguiente, cabe preguntarse: ¿cómo ocurre un cambio de paradigma? Resulta conveniente resaltar que un paradigma puede tener tanto un alcance extendido como limitado; puede afectar a varios campos de la ciencia como solo a un subconjunto de un campo.

El proceso mediante el cual se lleva a cabo la sustitución de un paradigma por otro es paulatino y conflictivo, genera tanto reacciones a favor como en contra. Existen profesionales tan arraigados a la tradición que están dispuestos a dar su vida por defender un conjunto de creencias en específico, como también otro conjunto extasiados por la emergencia de una nueva forma de hacer ciencia que están dispuestos a desechar totalmente las antiguas posturas por unas nuevas y emocionantes. Centrándonos en cómo sucede el cambio en sí, este se lleva a cabo por el tránsito de un paradigma a otro que no es más que el paso de un conjunto de creencias dado —que no es capaz de superar las anomalías— a otro que haya superado las anomalías presentes en el anterior; pero esto no es del todo cierto. El tránsito de un paradigma a otro se da por la presencia de una situación de crisis prolongada en el tiempo: estas situaciones de crisis son dadas por anomalías (Kuhn 2011, p. 27). Estas son dilemas o acertijos a los cuales el paradigma no puede hacer frente, es decir, que es incapaz de resolverlos. Comúnmente, antes de dar por perdido un paradigma surgen diversas variaciones y modificaciones en el mismo para poder asimilar la anomalía. Esta situación en sí no es grave, siempre y cuando no perdure en el tiempo o su resolución derive en un ajuste que provoque anomalías aún más complejas. Si un paradigma, pasado un largo periodo de tiempo, sigue sin ser capaz de superar la anomalía o conseguir un ajuste que pueda evitarla, empiezan a surgir posturas alternativas, teorías distintas, nuevas dinámicas, herramientas y métodos más allá del paradigma actual. Estos son momentos extremadamente creativos y suelen ser llevados a cabo por novatos dentro del campo de la ciencia, puesto que los veteranos (criados dentro del paradigma y aferrados al mismo) tienden a confiar en que el paradigma superará la anomalía; estos momentos de crisis e innovación son llamados ciencia extraordinaria (Kuhn 2011 , pp. 157-192).

Entre las diversas posturas alternativas (al actual paradigma) habrá una que se alce sobre las demás, primero por superar las anomalías, segundo por ser la que más partidarios posea. Esta será desarrollada por un conjunto de fieles, los cuales estructurarán poco a poco los nuevos cimientos del futuro paradigma e intentarán atraer a más científicos a formar parte de esta visión. El tránsito suele ser algo generacional, tanto por educación como por el desarrollo de la ciencia normal y una anomalía escapa de lo anteriormente mencionado de tal manera que aquellos científicos que se desarrollaron y fueron educados dentro de un mismo paradigma (el cual enfrenta la anomalía) difícilmente podrán

ser capaces de burlarlo —por recaer meramente en lo que el paradigma postula mientras que la solución se halla fuera del mismo—: son los jóvenes científicos educados en la crisis los cuales se aventurarán a plantear teorías novedosas y atrevidas para superarla. Una vez que sucede el nuevo paradigma, es adoptado por toda la comunidad científica (o la gran mayoría) y trae consigo diversas consecuencias: en primer lugar, un cambio enorme en lo que respecta a la ciencia, ya que una vez establecido el nuevo paradigma implica el descarte de las antiguas teorías, metodologías, conocimientos y herramientas; en resumen, de varias o todas las creencias del antiguo paradigma. A su vez, supone una reescritura de la historia de la ciencia, así como de los manuales educativos —debemos tener en cuenta que los científicos se educan según el conjunto de creencias del paradigma en el que se introducen—. Esto es extremadamente importante dado que al redactar de nuevo los manuales y la historia de la ciencia no se hace añadiendo todo lo sucedido en el paradigma anterior y cómo fue el proceso de transición, sino como una abstracción de lo sucedido, descartando la gran mayoría de acontecimientos pasados que no fueron exitosos. Esto resulta en una carente historia de hechos triunfales que se suceden uno tras otro, situación que también ocurre con los manuales donde realmente tiene más sentido que se realice. Sería contradictorio enseñar metodologías y teorías ya desestimadas.

Acabamos de toparnos con uno de los primeros mitos y trastornos en el desarrollo de la ciencia: en principio la parcialidad de su historia y como consecuencia el conocimiento limitado que tenemos de la misma. Es de vital importancia comprender que en estos casos al narrar la historia de la ciencia no existen realmente criterios objetivos para estructurarla de esta manera —solo sería justificable dentro de ciertas dinámicas especializadas—. Por lo tanto, o existen motivos externos a la ciencia que determinen esta decisión o un conjunto de autoridades científicas (no tienen porque que serlo necesariamente, puede ser cualquier tipo de individuo o individuos con gran influencia en la ciencia, pero ajenos a la misma) ha decidido restringir la información de la que disponía de forma arbitraria.

Ya hemos establecido el marco que nos servirá como depurador de aquellos resquicios ficticios que existen acerca de la ciencia y el que revelará las manipulaciones que pueden darse en la misma; con ello ahora enumeraremos algunos mitos que podemos dismantelar a partir de lo anterior:

- **La ciencia progresa:** de base esta concepción carece de sustento alguno. Para afirmar que existe un progreso en la ciencia debemos tener un punto de llegada, un objetivo por cumplir. Podríamos pensar que (1) este objetivo es alcanzar la verdad o (2) lograr establecer un paradigma científico que abarque más soluciones y que no recaiga en anomalías. Sin embargo, cualquiera que sepa un poco de filosofía de la ciencia sabe que (1) es muy atrevida y (2) peca completamente de utópica —por ser lejana al funcionamiento de la práctica científica—. Si las razones anteriores no son suficientes basta con observar el desarrollo científico para darse cuenta de que realmente el paso de un paradigma a otro no implica necesariamente un avance (Kuhn 2011, pp. 268-289, 313-315). De hecho, al contrario, a menudo es incluso una regresión (Kuhn 2011, pp. 268-289). El paso de un paradigma a otro viene cargado de una gran cantidad de nuevos acertijos por resolver e incluso suele traer de nuevo a discusión cuestiones ya resueltas por el anterior. Con la llegada del nuevo paradigma este es capaz de explicar menos fenómenos que su antecesor, no es hasta que es desarrollado de forma extendida que puede considerarse un avance —esto claro hasta que surja una anomalía—. No se puede olvidar que la llegada de un nuevo paradigma implica descartar teorías, ejercicios, metodologías y herramientas pertenecientes al paradigma anterior; a su vez, debemos considerar que si no hay un paradigma nuevo para suplir al que está en crisis, este permanecerá en ese estado hasta la aparición de un digno sucesor. Con todo ello, la ciencia se ve envuelta en un ciclo constante donde: 1) Se establece un paradigma, 2) surge una anomalía y 3) se sustituye por un nuevo paradigma. Dados estos tres pasos inicia el bucle a partir del tercer paso, se vuelve al segundo, del segundo al tercero y así constantemente. Concluyendo con este punto como máximo, podemos establecer que la ciencia está en constante desarrollo, pero esto no tiene por qué significar que la ciencia progrese de alguna forma.
- **Elección entre paradigmas:** anteriormente pasamos por alto discretamente este tema y no sin razón, pues este es uno de los principales puntos en los que nuestro esquema se ajusta a la realidad científica. No solo cuando se produce la transición de un paradigma a otro, sino también cuando se selecciona un nuevo paradigma de entre las opciones posibles. Por lo tanto, hay dos

momentos de decisión en los que encontramos criterios de elección que no son objetivamente científicos. En primer lugar, nos centraremos en la decisión de proponer un nuevo paradigma y posteriormente en la decisión de pasar de un paradigma a otro:

Como explicamos anteriormente, cuando un paradigma entra en una crisis prolongada de la cual no consigue salir, empiezan a proponerse nuevas teorías fuera del paradigma que logren resolver la anomalía a la cual este no puede hacer frente. De todas aquellas teorías emergentes solo una será el futuro paradigma que sustituirá al actual; pero para que eso suceda, primero deberá ser elegido. Se cree que una teoría se elige sobre otras en cuanto puede superar más pruebas que las demás y puede explicar y solucionar más anomalías que sus contendientes. En la lógica científica sería lo más coherente, pero en la práctica científica la realidad es distinta. Hemos de comprender que cuando queremos elegir entre teorías, primero los partidarios de una teoría en específico suelen ser acérrimos defensores de estas y han de recibir argumentos de mucho peso para desligarse de su postura; segundo, las teorías científicas distintas contienen tanto un lenguaje como una práctica completamente diferente, entendible solo para aquellos partícipes de esta. Como resultado, en algún momento, los defensores de su teoría se ven obligados a defenderla con sus argumentos más fuertes, los cuales indudablemente estarán cargados de un montón de lenguaje técnico y prácticas de la misma teoría. Ahora imaginemos la situación: dos científicos discutiendo y defendiendo sus posiciones a través de argumentos que solo entienden cada uno por la configuración de sus propias teorías y vocabulario específico. De tal manera que los científicos están obligados a discutir sin entenderse y a partir de ahí esperar convencerse el uno a otro en una batalla entre un robot y un bárbaro. Como resultado, ninguno de ellos puede esperar probar sus argumentos, al igual que ninguna de las dos partes concederá todas las suposiciones no empíricas que necesita la otra para desarrollar su argumento. Entonces como dirá Kuhn “la competencia entre paradigmas no es el tipo de batalla que pueda resolverse por medio de pruebas” (Kuhn 2011, p. 230). Gran parte de la adopción de un nuevo paradigma viene acompañada de la retórica y la visión estética que pueden aportar del mismo para atraer a esta primera oleada que estructurará la teoría. En cuanto una teoría es capaz de superar la anomalía, lo que determinará si es elegida sobre sus

otros contendientes es más la sofística y el marketing que se haga de la misma antes que los criterios científicos objetivos. Esto no es más que otro criterio arbitrario dentro de la disciplina científica, poco tienen que aportar los criterios epistémicos y objetivos en cuanto predomine lo sublime y lo bello para promocionar el nuevo paradigma; en cuanto se agota el alcance de nuestros filtros epistémicos, la estética toma el rol de legislador en la decisión (o mejor dicho, de promotor).

En cuanto a este tema seremos más breves: el tránsito que existe entre una teoría y otra es extenso por varias razones. Primero, por la conocida resistencia de los miembros educados y practicantes dentro del paradigma; segundo, por la juventud y poca experiencia que se percibe en los miembros de la nueva teoría; y tercero, por lo mencionado en cuanto a la elección de paradigmas. Lo que explicábamos en la situación anterior se da también en esta; resistencia ante las nuevas propuestas, dificultades a la hora de comunicarse entre miembros de posturas distintas, falta de criterios para establecer qué teoría es más conveniente. La diferencia crucial entre ambos momentos es: si en las situaciones señaladas anteriormente se termina tomando una decisión, pese a ser por razones subjetivas más que objetivas, en este caso sucede lo contrario, no se llega a tomar una decisión en sí, lo más común es que la transición no sea una cuestión de decisión, en cambio es una transición generacional; hasta que no muere la generación que defiende el paradigma actual no se establece uno nuevo.

#### 4. Las motivaciones de la Ciencia

Cuando hablamos de manipulación en la ciencia hacemos hincapié en la falta de criterios objetivos al discernir en las situaciones que nuestro esquema visibiliza. Por consiguiente, existen motivos ajenos que impulsan estas decisiones que van más allá de la ciencia y su comunidad. No debemos perder de vista que la ciencia como institución se ve envuelta en el entramado del desarrollo histórico (Khun 2011). Con ello, establece relaciones —no siempre por voluntad de la comunidad científica— con diversas instituciones y entidades que ejercerán su influencia sobre la misma y viceversa. Entenderemos aquí que la ciencia funciona también como una institución social (Latour 1992) al igual que la religión, los hospitales o la familia, donde como en toda institución afecta a todos los individuos y estos a su vez intervienen en la



misma. En este caso, no estamos ofreciendo una definición ni articulando un concepto que encierre a la ciencia, lo que queremos ilustrar es que la ciencia, la práctica y el conocimiento científicos son atravesados igualmente por todo el entramado social por el que atraviesa (y es atravesado) cualquier dinámica en la que opere el ser humano, y como tal se ve influido por una multiplicidad de factores, campos e instituciones. Nos referimos a ella como institución en tanto la ciencia implica una organización, unas ciertas normas y leyes dentro del desarrollo científico para con su fin. El científico será otro miembro más de la tesitura social que ejerce un oficio. Es obvio que el principal factor de intervención es el contexto histórico (Kuhn 2011) pero esta sería una lectura muy superficial en nuestra labor. Nos interesa más nombrar factores específicos que a lo largo de la historia hayan desempeñado un papel fundamental en el desarrollo del ámbito científico y que aún hoy en día siguen teniendo gran influencia.

El primero de ellos es el entramado político, tanto los estados como sus instituciones y la sociedad política son de vital importancia en el desarrollo de la ciencia. La política ha influido de sobremano siempre en el ámbito científico, impulsando ciertos desarrollos, coartando otros, financiando proyectos como también exigiéndolos. La autoridad política es una de las principales asociaciones que tiene la ciencia (Latour 1992) (Longino 1990). Basta con observar cómo el desarrollo científico viene determinado tanto por el proyecto político de turno como por la situación del estado. Ejemplos como iniciativas contra el cambio climático (auspiciando el desarrollo de proyectos de este tipo ante otros), situaciones de guerra (la presencia de la ciencia en la guerra es indudable. Los equilibrios geopolíticos y geoestratégicos del ejército han cambiado en los últimos años debido a la influencia de la ciencia; un claro ejemplo de ello son las carreras armamentísticas) buscar impulso tecnológico en algún mercado o simplemente la aceptación o prohibición de ciertos proyectos o dinámicas son aspectos que la política impone sobre la ciencia. El desarrollo sociopolítico promueve la investigación en algunos campos sobre otros tan solo por meros intereses.

Otro factor de inmensa relevancia es el social; la sociedad recibe constantemente los efectos del desarrollo científico. Nadie pone en duda el poder que tiene el factor social y su reacción ante proyectos científicos, y el recibimiento que se tenga de los mismos marcará el desarrollo que se llevará a cabo. La ciencia como institución ha de cumplir ciertos marcos (morales, políticos) que le exige la sociedad y con ello delimitar qué acciones puede llevar a cabo dentro de su desarrollo. En ella tam-

bién pueden verse representados intereses de la comunidad social por apoyar ciertos proyectos o campos de investigación sobre otros y con ello proporcionando mayor apoyo tanto económico como mediático.

Existe un factor más que interviene y motiva a la ciencia. Pese a ser el último tratado, puede que sea el de mayor relevancia hoy en día: este es el económico, el mercado, las empresas, la demanda y las exigencias que motivan en gran medida el desarrollo científico (Mirowski 2001). Esto no es extraño, la financiación no sale de debajo de las rocas. Esto conlleva consecuencias para la comunidad científica; deberán realizar los proyectos que se decidan financiar —ya sea por grandes empresas, el estado o algún millonario filántropo de turno— y con ello limitando los campos o proyectos que se puedan llevar a cabo. La propia ciencia también tiene motivaciones económicas: no se producen ni se realizan proyectos que no puedan proporcionar ganancia o generar suficiente interés para financiarlos; todas las decisiones toman en cuenta estos factores. Una situación no muy común, pero existente, es el poder de decisión de los propietarios del capital en las investigaciones, pudiendo usar su poder económico para modificar o redireccionar el proyecto que se esté llevando a cabo a su antojo.

Estos tres elementos que acabamos de nombrar —lo económico, lo social y la política— son los principales influyentes a la hora de la toma de decisiones dentro de la ciencia, más allá de los criterios de esta; con ello ya podemos comprender cómo se pueden producir decisiones de forma arbitraria a favor de un interés ajeno al desarrollo científico. De ahí se sigue que las manipulaciones del conocimiento científico pueden haber sido motivadas por intereses tanto económicos como sociales y políticos. Ahora que completamos el esquema, nos hemos hecho de una herramienta que nos permite analizar ciertos aspectos dentro de la ciencia y consigo explicitar los trastocamientos y manipulaciones que podemos hallar en la misma. Con lo expuesto, queremos resaltar que ha sido una crítica en sentido laxo. No buscábamos arremeter ni atacar al conjunto científico y mucho menos prescribir cómo debería desenvolverse la práctica científica, sino comprender y analizar su funcionamiento. Para así obtener una visión más cercana al funcionamiento real de la dinámica científica y romper ciertos mitos que existen entre aquellos menos familiarizados con este campo. La ciencia en su conjunto es un ámbito que nos afecta de diversas maneras hoy por hoy y es necesario comprenderla adecuadamente para poder entender la estructura de la sociedad actual. Con esto dicho, esperamos ofrecer una herramien-

ta que permita filtrar nociones erróneas para obtener una perspectiva completa y clara de la ciencia en nuestros días.

## Referencias

- Aristóteles. (1994). *Metafísica*. Madrid. Gredos.
- Baudelaire, C. (1857). *Les Fleurs du mal: Les phares*. Alenzón. Auguste Poulet-Malassis.
- Comte, A. (1980). *Discurso sobre el espíritu positivo*. Versión y prólogo de Julián Marías. Madrid. Alianza Ed.
- Darcy, J. (2007). *Introducción a la epistemología contemporánea*. Madrid. Tecnos.
- Fricker, M. (2017). *Injusticia epistémica*. Barcelona. Herder.
- Fundación Gustavo Bueno. (2017). <http://fgbueno.es>. (s. f.). Enciclopedias y diccionarios: Sistema geocéntrico y heliocéntrico del mundo. <https://www.filosofia.org/enc/ros/sis4.htm>.
- Gettier, E. (1963). "Is Justified True Belief Knowledge?". *Oxford Analysis*. 23 (6) , pp. 121–123.
- Hempel, C. (1996). *La explicación científica: Estudios sobre la filosofía de la ciencia*. Barcelona. Paidós.
- Hintikka, J. (2007). *Socratic Epistemology: Explorations of Knowledge-Seeking by Questioning*. Cambridge. Cambridge University Press.
- Kuhn, T. (2011). *La estructura de las revoluciones científicas*. Ciudad de México. Fondo de Cultura Económica de México.
- Latour, B. (1992). *Ciencia en acción: Como seguir a los científicos e ingenieros a través de la sociedad*. Barcelona. Labor.
- Longino, H. (1990). *Science as Social Knowledge*. New Jersey. Princeton university press.
- Mirowski, P. y Sent, E.-M. (2001). *Science Bought and Sold*. Chicago. University of Chicago press.

Platón. (2021). *Teeteto*. Diálogos v: · Parménides · Teeteto · Sofista · Político. Madrid. Gredos. 137-317.

Young, J. (2003). La teoría de la verdad como coherencia. Victoria. *Discusiones Filosóficas Año 4 N°7* , pp. 110-120.

# Reflexiones desde un enfoque biosemiótico peirceano sobre algunos procesos de mimetismo biológico.

Loreto Paniagua-Valdebenito<sup>a</sup>

## Resumen

La biosemiótica es un campo interdisciplinario donde las ciencias y las humanidades convergen de forma complementaria. Una de las áreas más fértiles ha sido la relación entre las ciencias biológicas y la semiótica peirceana, ya que esta última propone una perspectiva novedosa sobre la comunicación no humana y el papel del signo como propiedad relacional de los procesos de semiosis presentes en la naturaleza. La propuesta tardía del filósofo C.S. Peirce, especialmente su división en diferentes interpretantes, junto con algunos elementos de su epistemología como su máxima pragmática y su sistema categorial, configuran la comprensión de cómo las consecuencias prácticas de algunos fenómenos interespecie se constituyen como cambios en los hábitos de acción con implicaciones en los comportamientos o acciones de diversas formas de vida animal. De esta manera, se revisan algunos casos de mimetismo biológico desde la tríada semiótica peirceana, resaltando la iconicidad y las implicaciones prácticas que realizan algunos animales e insectos.

**Palabras clave:** Pragmaticismo, ícono, comunicación, información, mimetismo.

## Abstract

---

<sup>a</sup>Investigadora predoctoral ECyT, Universidad de Salamanca (España), Docente Universidad de Valparaíso.

Contacto: loreto.paniagua@usal.es

Biosemiotics is an interdisciplinary field where sciences and humanities converge in a complementary way. One of the most fertile areas has been the relationship between biological sciences and Peircean semiotics since the latter proposes a novel perspective on non-human communication and the role of the sign as a relational property of the processes of semiosis present in nature. The late proposal of the philosopher C. S. Peirce, especially his division into different interpretants, together with some elements of his epistemology such as his pragmatic maxim and his categorial system, shape the understanding of how the practical consequences of some Inter-species phenomena are constituted as changes in habits of action with implications in the behaviors or actions of various forms of animal life. In this way, some cases of biological mimicry are reviewed from the Peircean semiotic triad, highlighting the iconicity and practical implications that some animals and insects perform.

**Keywords:** Pragmatism, icon, communication, information, mimicry.

## 1. Introducción

La semiótica peirceana es la propuesta que actualmente se consolida para el trabajo en biosemiótica (Kull et al. 2011; Barbieri 2009). Principalmente porque la filosofía peirceana proporciona algunas relaciones y elementos que son necesarios para comprender algunos procesos en el mundo natural superando en primer lugar la concepción subjetivista del conocimiento heredada de la modernidad, en específico el dualismo mente cuerpo, y así mismo la pérdida de continuidad entre aquellos eventos mentales y otros de índole material. Y en segundo lugar, porque la supera los marcos de comprensión de la comunicación aportados por la semiología. Así, la biosemiótica como disciplina que intenta comprender los modos comunicacionales entre diversas formas de vida y sus implicaciones en la evolución, como también las formas relacionales entre especies, necesita un marco teórico que no sólo considere los procesos de semiosis, sino que asimismo intente explicar cómo estos se encuadran en una formulación epistemológica más amplia (Romanini 2014); la propuesta del filósofo Charles Sanders Peirce aporta ambos beneficios de manera armoniosa.

Este artículo expone ciertos componentes de la epistemología peirceana, específicamente los aportes de las categorías y la máxima pragmática, y cómo estos elementos teóricos complementan la comprensión del carácter práctico de la semiótica tardía del filósofo. Paralelamente, se intenta relacionar la teoría de los signos como una perspectiva fundamental para la biosemiótica, frente a otras aportaciones en este campo que no se ajustan adecuadamente a las exigencias de los fenómenos

biológicos en su generalidad. Finalmente, se aplica la estructura expuesta por Timo Maran (2017) para un breve análisis semiótico a casos de mimetismo biológico determinados.

## 2. Algunas aportaciones de la filosofía de C.S. Peirce como marco teórico de la biosemiótica

Los elementos centrales de la propuesta pragmaticista son las categorías y la máxima pragmática. Si bien ambos aportes teóricos se originan y consolidan en períodos diferentes, hacia el final de la vida del filósofo estas estructuras convergen, posibilitando un nuevo marco epistemológico. Lo interesante de esto es que este marco teórico es apropiado hoy para el estudio de una amplia gama de disciplinas, además de proporcionar una base común para una comprensión integral del conocimiento.

El pragmaticismo es una herramienta para hacer nuestra mente receptiva a la evidencia, esto puede entenderse como una propuesta de configuración general donde se expresa el realismo peirceano para la comprensión de los fenómenos. En el caso de la máxima pragmática, como componente fundamental de esta propuesta, existen algunos conceptos claves como: irritación de la duda, creencia (CP 5.394) y hábito de acción (CP 5.400), los cuales se articulan dando forma a nuestra comprensión de los fenómenos. Esto, acompañado de la propuesta categorial del filósofo (que se compone de primeridad, segundidad y terceridad), da unidad al conocimiento para la comprensión de fenómenos a través de la ciencia. En biosemiótica, la importancia de las categorías se expresa de la siguiente manera:

*“The fundamental categories should provide a means of unifying all reality into an intelligible continuum. An explanation of the phenomenon of life should also be an explanation of the phenomenon of intelligence and meaning since every genuine triadic relationship presupposes meaning”* [Las categorías fundamentales deberían proporcionar un medio de unificar toda la realidad en un continuo inteligible. Una explicación del fenómeno de la vida debería ser también una explicación del fenómeno de la inteligencia y del significado ya que toda relación triádica auténtica presupone un significado] (CP 1.345 en Romanini y Fernández 2014, p. 8).

Una de las propuestas de este artículo es que la semiótica de Peirce debe necesariamente pensarse en conjunto con los elementos del marco epistemológico del pragmatismo. Si la semiótica se separa de estos recursos epistémicos no pasa de ser una herramienta de decodificación que sólo se utiliza como método de traducción sin la significación y alcance práctico que tiene hoy. Así, la semiótica peirceana alcanza su más alta expresión acompañada de la máxima pragmática y las categorías. Es de esta manera que el carácter instrumental de la semiótica se proyecta hacia la conformación de un modelo para una lectura no antropocéntrica de la transferencia de información y la comunicación.

Por el contrario, generalmente desde la filosofía de la biología se entiende la comunicación en los diversos niveles de la vida a partir de un modelo de traspaso de señales, como muy bien lo relata Godfrey-Smith a través del análisis de algunos modelos comunicacionales como el modelo de la información de Claude Shannon y el modelo de David Lewis (Godfrey-Smith 2014). Para esta investigación, ese enfoque es insuficiente por al menos tres razones: 1. solo es aplicable reduciendo los fenómenos comunicacionales a dualismos emisor-receptor o comunicador-audiencia, lo cual conlleva una serie de supuestos y problemas. 2. ambos modelos entienden las señales como “algo” que se traspasa, lo cual filosóficamente implica un compromiso ontológico que otras perspectivas no están dispuestas a asumir. 3. no da la suficiente importancia a las consecuencias prácticas de la relación comunicativa, por lo cual no es capaz de comprender de forma amplia los eventos comunicacionales. Sin embargo, estas consideraciones son suficientes para otro estudio, por lo cual no serán tratadas de forma exhaustiva en este artículo.

Ahora bien, retomando los aportes del pragmatismo a la semiótica peirceana, según Atkin (2022) existe una relación directa entre la semiótica tardía de Peirce y la máxima pragmática presentada décadas antes. Los tipos de claridad (familiaridad, análisis lógico, análisis pragmático) expuestos en el texto *How to make our Ideas Clear* (CP 5.388-410, p. 1878) potencian las distinciones necesarias. En la explicación del tercero de ellos (análisis pragmático) es donde Peirce expresa una de las definiciones que luego será considerada como el interpretante final, el cual es clave en la construcción posterior de su semiótica ya que el carácter pragmático depende en gran medida de esta definición:

*“It appears then that the rule for attaining the third grade of clearness of apprehension is as follows: Consider what effects which might conceivably have practical bearings we*



*conceive the object of our conception to have. Then our conception of these effects is the whole of our conception of the object.*” [Parece pues que la regla para alcanzar el tercer grado de claridad de aprehensión es la siguiente: Considérense los efectos que en nuestra opinión puede tener el objeto de nuestra concepción. Entonces nuestra concepción de estos efectos es la totalidad de nuestra concepción del objeto]. (CP 5.402)

La semiótica que Peirce propone, desde sus propuestas más básicas hasta las más complejas, busca resguardar la función del interpretante no como un sujeto que interpreta (lo cual sería mantener la perspectiva dualista y subjetivista cartesiana), sino más bien como aquellas posibles consecuencias prácticas de la relación entre el signo (representamen) y el objeto. Cada una de estas consecuencias puede funcionar posteriormente como un signo y así producir una cadena de semiosis. Esta estructura es revolucionaria por sí sola porque muestra la continuidad de la comunicación y sus alcances.

Sin embargo, esto sumado al desarrollo de su sistema categorial proporciona un modelo comunicativo que es capaz de adaptarse a una gama más amplia de eventos que otras teorías de la comunicación y la transferencia de información. Identificar la funcionalidad pragmática de la semiótica peirceana y de su estructura triádica es un giro que rompe con las estructuras de la comunicación como fenómeno exclusivamente humano, facilitando que esta propuesta pueda ser el marco de comprensiones nuevas sobre la comunicación no lingüística e interespecie, entre otras modalidades posibles.

Charles Morris (1971), por su parte, destacó la practicidad y la naturaleza problemática de la propuesta peirceana para comprender otras formas de comunicación presentes en animales de otras especies. En sus *Writings on the General Theory of Signs* lo expresa de la siguiente manera:

*“His Classification of signs, his refusal to separate completely animal and human sign-processes, his often penetrating remarks on linguistic categories, his application of semiotic to the problems of logic and philosophy, and the general acumen of his observations and distinctions make his work in semiotic a source of Simulation that has few equals in the history of this field.”* [Su Clasificación de signos, su negativa a separar completamente los procesos de signos animales y

humanos, sus comentarios a menudo penetrantes sobre las categorías lingüísticas, su aplicación de la semiótica a los problemas de la lógica y la filosofía, y la perspicacia general de sus observaciones y distinciones hacen que su obra en semiótica una fuente de simulación que tiene pocos iguales en la historia de este campo]. (p. 340)

La propuesta peirceana también es considerada por Short, pero más de treinta años después; con ello se pueden proyectar con mayor claridad los beneficios de la perspectiva no antropocéntrica de la comunicación desde la consolidación y comprensión de la transferencia de información desde múltiples rutas en la naturaleza hasta la comunicación intencionada en otras especies. Short (2007) explica claramente en el siguiente extracto cómo Peirce está en desacuerdo con una tradición histórica la cual propone una oposición entre naturaleza e intencionalidad:

*“In Augustine’s time as in ours, nature was ambiguously conceived of as one part of any of various oppositions. Four centuries earlier, in De rerum natura Lucretius opposed nature to purpose. That would seem to be the opposition that Augustine had in mind when he said that signa naturalia involve no intention to signify. Smoke does not occur in order to draw attention to fire, but animal signals are meant to elicit specific sorts of response. Augustine thus redrew the fundamental division of signs: instead of being between the natural and the conventional, it was between the natural and the purposeful, with conventional signs being a subset of those signs that occur for a purpose. Peirce’s idea of ‘legisigns’ and their replicas is exactly Augustine’s of signa data and it applies to the same range of examples. Peirce, however, did not conceive of nature as that to which the purposeful is opposed.”* [En la época de Agustín, como en la nuestra, la naturaleza era concebida ambiguamente como una parte de varias oposiciones. Cuatro siglos antes, en *De rerum natura*, Lucrecio opuso la naturaleza al propósito. Esa parecería ser la oposición que Agustín tenía en mente cuando dijo que la *signa naturalia* no implica ninguna intención de significar. El humo no se produce para llamar la atención sobre el fuego, pero las señales de los animales pretenden provocar tipos específicos de respuesta. Así, Agustín volvió a trazar la división fundamental de los signos: en lugar de

estar entre lo natural y lo convencional, estaba entre lo natural y lo intencionado, siendo los signos convencionales un subconjunto de aquellos signos que ocurren con un propósito. La idea de Peirce sobre los “legisignos” y sus réplicas es exactamente la de Agustín sobre los *signa data* y se aplica al mismo rango de ejemplos. Peirce, sin embargo, no concibió la naturaleza como aquello a lo que se opone lo intencional]. (pp. 25-26)

Para Peirce, la continuidad sígnica está presente en todo proceso de conocimiento, y para el filósofo esto se manifiesta en uno de sus principios cosmogónicos fundamentales. Esta característica relevante es uno de los elementos clave para que la propuesta peirceana sea un aporte a los nuevos enfoques sobre el traspaso de información y así mismo de los eventos comunicativos en diversos niveles y formas de vida. Peirce no opone lo intencional a la naturaleza porque para él hay objetivos, propósitos, pero no aquellos pensados desde una perspectiva teleológica tradicional aristotélica que presupone intencionalidad racional o presuponiendo un sujeto que diseña inteligentemente, sino muy diferente a lo anterior, estos propósitos serían lo que hoy en día Daniel Dennett ha denominado con suma elegancia *enfoque de diseño* (Dennett 2024). Esta perspectiva no contraría el azar considerado en la teoría Darwiniana de la evolución, sino que da cuenta de procesos que se abren paso a través de posibilidades prácticas de la comunicación o el traspaso de información (como es el caso de los genes), y que el ser humano solo llega a desentrañar estos procesos realizando “ingeniería inversa”.

Dennett explica que hay un “conjunto de procesos que “hallan” y “rastrear” razones para que las cosas se dispongan de una determinada manera y no de otra” (2017, p. 44). Esto epistemológicamente no es trivial. Sin embargo, las “razones” que encuentra la evolución o los procesos biológicos, y las razones que construyen intencionalmente los humanos, no son el mismo tipo de proceso. Las razones humanas suelen estar primero ideadas o representadas en nuestras mentes; muy diferente a lo anterior, las razones de la evolución o de los procesos biológicos solo podemos conocerlas a través del estudio, desde desarticular y comprender dichos procesos de manera inversa, comprendiendo sus causas materiales, físico-químicas, causales, entre otras.

En esta ingeniería inversa aplicada a la materia viva suponemos intencionalidad porque estamos acostumbrados a un sujeto constructor y creador inteligente cuando se trata de artefactos o construcciones artificiales. No obstante, cuando se trata de mecanismos bioquímicos o

físicos no podemos hablar de intencionalidad *per se*, pero sí de convergencia, traspaso de información y azar en relación con un sustrato material. La semiótica pone de relieve estos propósitos en sus variadas manifestaciones fenoménicas.

Otro de los aportes clave de la propuesta peirceana es la división de tipos de interpretantes. En 1906, Peirce ya había extendido su noción de interpretante como sentimiento. Luego llama a estas extensiones conceptuales interpretantes emocionales, energéticos y lógicos. Short afirma que Peirce apenas menciona a interpretantes no humanos, pero es en una carta a Lady Welby de 1908 donde comenta su anhelo de construir una concepción semiótica más amplia que la que había desarrollado hasta ese momento (Short 2007). Ya la tricotomía de 1906 de los tipos interpretantes se basa en los efectos significativos mismos, esto es, en la disposición práctica que se evidencia en las relaciones semióticas. Según Lalor (1997), esto se refiere a la apreciación de Peirce de que el primer efecto de un signo es un sentimiento que produce el interpretante emocional. Mientras que el interpretante energético es aquel que realiza un esfuerzo mental o muscular como efecto del interpretante emocional que media entre el fenómeno y la acción. Esta estructura pragmática es relevante porque es posible aplicarla a las conductas de interpretantes no humanos (aunque cabe señalar que cuando Peirce construyó esta tricotomía no la extendió a animales de otras especies de forma explícita). Si asumimos que las consecuencias prácticas (interpretantes) en animales no humanos son en algunos casos emocionales, en otros energéticos, e incluso lógicos, también asumimos que estos animales realizan procesos inferenciales, lo cual es una puerta de entrada hacia la cognición animal como un proceso de nivel superior comunicacionalmente. En el siglo XXI, la capacidad sintiente en gran parte del reino animal ya no se pone en duda entre especialistas.

En los ejemplos de la siguiente sección se utilizará la tricotomía semiótica para analizar conductas como interpretantes. Por ejemplo, cuando un animal huye de algo determinado después de percibir un acto mimético icónico que entrega información de un posible peligro (e.g. el depredador de una *Heliconius numata* con el patrón de sus alas similar al de una *Malinea mneme* que evita cazarla dado que infiere que esta tendrá mal sabor). No es irrelevante utilizar el concepto de inferencia para referirse al rol del interpretante presentado en ciertos casos de semiosis. Este punto es otro de los aportes de la perspectiva peirceana, la cuestión cognitiva e inferencial, sin embargo, autores como Lorenzo Magnani y Woosuk Park han tratado este tema a profundidad.

El último componente de esta propuesta tricotómica es el interpretante lógico, el cual puede entenderse de la siguiente manera:

*“the real and living logical conclusion is that habit; the verbal formulation merely expresses it”* [la conclusión lógica real y viva es ese hábito; la formulación verbal simplemente lo expresa] (CP 5.491).

Si el interpretante lógico es el hábito, los animales de otras especies no están exentos de este tipo de relaciones. Así, es posible extrapolar esta estructura y proyectarla más allá de su expresión verbal o formal, centrándose en la generación de hábitos de acción como interpretante lógico inferencial sin necesidad de una construcción proposicional.

En el período filosófico posterior a 1906, a través del diálogo epistolar con Victoria Welby, Peirce desarrolla estructuras de mayor complejidad que resultan útiles para comprender los efectos prácticos de los procesos de mimetismo biológico en otras especies. Dentro de la etapa posterior a 1909 en la obra de Peirce podemos encontrar una segunda propuesta semiótica. Esta tricotomía tiene la siguiente estructura: interpretante inmediato, dinámico e interpretante final. Algunos académicos sostienen que las propuestas de 1906 y 1909 son completamente diferentes entre sí, mientras que otros postulan que la última es una extensión de la primera (Lalor 1997).

Con respecto a las categorías, Romanini (2014) en su texto *Semeiosis as a Living Process* expone que uno de los aportes intelectuales de Lady Victoria Welby es resaltar la importancia de las categorías y cómo estas se relacionan directamente con las clases de signos. En *To Lady Welby On Signs and the Category* el filósofo explica la relación entre las categorías y su semiótica madura (CP 8.327) lo que sirve posteriormente para considerar este sistema categorial como un elemento indispensable para el desarrollo de la perspectiva biosemiótica porque es esta relación epistemológica la que relaciona la semiótica con el mundo sin separarla de lo que Peirce entendía por lógica.

Existe un consenso de que la semiótica peirceana evoluciona a través de las etapas dentro de su obra; estas se diferencian en al menos tres épocas siguiendo la propuesta de Albert Atkin (2022). En estas etapas los elementos fundamentales que componen la teoría de los signos se complejizan y amplían la perspectiva semiótica. Sin embargo, más allá de las complejidades documentales asociadas a seguir una posible definición del concepto de signo o la definición general de la semiótica, hoy

vemos más claramente gracias a diversos esfuerzos, que el factor temporal en el progreso y evolución de la obra de Peirce no es baladí, y la comprensión de este desarrollo implica claridad conceptual al intentar entender en profundidad la estructura arquitectónica de su filosofía. De este modo, la semiótica tardía está más claramente relacionada con el sistema categorial y la máxima pragmática, pero el hecho más relevante es que tiene influencia del conocimiento de la época. Peirce y Welby comparten el paradigma biológico inaugurado por Darwin aplicándolo a las diferentes ciencias y a la práctica social (Petrilli 2023). Esta es la posible causa de la extensión naturalista de la semiótica peirceana. Peirce no estaba ajeno a una visión Darwiniana de la evolución, es esta la que posteriormente Charles Morris desarrolla conceptualmente como uno de los puntos claves de su filosofía, a saber, la atribución de la semiosis a los organismos vivos, propuesta que lleva a Thomas Sebeok a inaugurar la biosemiótica como un campo independiente (Petrilli & Ponzio 2005) y a otras filósofas como Donna Haraway a utilizarla como contexto para su teoría de la comunicación animal (Haraway 1989). Con lo anterior se hace comprensible y coherente por qué Peirce siempre tuvo presente el azar como vehículo constitutivo del que depende todo proceso (MS 875).

Por último, otra aportación de la semiótica peirceana frente a otras propuestas teóricas es que difiere de la semiología de Ferdinand Saussure, específicamente en su idea de signo y no sólo en el número de elementos combinados. Más bien, lo relevante es la forma en que se entienden los componentes de ambas teorías. Para Short (2007), la diferencia es clara y radica en que para Saussure el signo es una entidad compuesta y para Peirce una propiedad relacional. Esta propiedad relacional está bien explicada en el siguiente texto de Merrell:

“el representamen no puede existir como tal sin un vínculo con su respectivo ‘objeto semiótico’ y su interpretante (que también son signos) y el símbolo contiene o depende de procesos icónicos e indexicales, al igual que el legisigno sobre las propiedades de cualisigno y sinsigno” (1997, p. 142).

La explicación de Merrell tiene a la base la estructura categorial del pragmaticismo. En esta, la terceridad también contiene o depende de la primeridad y la segundidad. De este modo, para llegar a una modelización completamente abstracta de un fenómeno del mundo, como por ejemplo un modelo matemático, es necesario el fenómeno (primeridad) y su relación con las consecuencias prácticas (segundidad) que produce.

La semiótica como aquella disciplina que se encarga de las relaciones entre el signo, el objeto y el interpretante, entendida como una forma de conocimiento, es capaz de proveer una aproximación a los fenómenos dinámicos y necesariamente relacionales del mundo natural.

### 3. El ícono para la comprensión del mimetismo biológico

Para llevar a cabo el análisis biosemiótico de esta sección es fundamental comprender la puesta en valor del ícono como elemento semiótico primario. Si bien el desarrollo de la semiótica de Peirce, presente en algunos extractos de las cartas enviadas a Welby, es mucho más complejo que lo que aquí se presenta para el análisis de los actos miméticos, no es necesario profundizar de manera exhaustiva en los diversos elementos expuestos por el filósofo. Sin embargo, esto no pretende separar la iconicidad del contexto semiótico general; más bien desea abrir la discusión a posibles nuevas perspectivas relacionales entre el mimetismo biológico y los conceptos principales de esta investigación.

Siguiendo la lectura de *Peirce on the Index and Indexical Reference*, Albert Atkin expone que Peirce en la etapa final de su obra da cuenta de la importancia de la implicación icónica destacando la relevancia práctica del ícono. En su texto Atkin afirma que el filósofo explica cómo la presencia de un ícono es esencial para la transferencia de información (CP 5.75 en Atkin 2005) y que es una característica fundamental de un ícono involucrado en un proceso de semiosis transmitir información.

Es en este contexto que Peirce construye la división de más de un objeto (inmediato y dinámico) y tres posibles interpretantes (inmediato, dinámico y final) lo que ayuda a comprender el carácter icónico, por ejemplo, del mimetismo biológico y cómo el interpretante (dentro de una mimesis funcional) es generalmente inmediato dado el objetivo principal del imitador que cumple el rol de representamen, “engañar” para favorecer algún tipo de proceso vital. Por su parte, el interpretante desarrolla la identificación de las consecuencias prácticas del proceso de semiosis desde una impresión perceptual primaria. En este caso, el signo y su relación con el objeto serían un medio para la generación de la impresión sensible que desencadena consecuencias prácticas.

Autores como Lalor y Romanini estudian este período filosófico encontrando algunas peculiaridades dignas de destacar. Por ejemplo, Peirce expresa la intención de abandonar la palabra “signo” por la palabra “medio”, la cual no estaría contaminada con confusiones históricas (EP

II 544). Por tanto, el signo para Peirce es un medio de comunicación, un vehículo. A su vez, los demás elementos (objeto e interpretante) también serían medio de comunicación. Peirce reflexiona estos vínculos desde dos concepciones fundamentales: la primera es la de “cadenas de semiosis” que se generan por la capacidad significativa de los tres componentes de la semiótica. En segundo lugar, la concepción cosmogónica de “continuidad” en los procesos de semiosis, que es desarrollada ontológicamente mediante la concepción del Sinejismo peirceano que nace desde su sistema categorial (CP 6.202).

Entonces, se entenderá que el acto mimético se fundamenta en la iconicidad entre el imitador y el modelo; es el ícono el que posee la misma cualidad de significado (Houser 2014). En el caso de conductas o características miméticas, el ícono permite variar las vías comunicativas, por lo que el interpretante puede ser inmediato (en el sentido de que el acto mimético es efectivo y confunde las propiedades del ser vivo con su mimetismo) o puede ser un interpretante dinámico cuando, por ejemplo, un depredador puede reconocer un mimetismo determinado y cambiar su comportamiento (hábito de acción) dependiendo de si la situación lo favorece (por ejemplo, si un depredador se entera de que el pulpo *Thaumoctopus mimicus* comete el acto mimético y distingue esta acción permitiendo cazar al pulpo), o un interpretante final cuando de algún modo son percibidas todas las posibilidades prácticas entre el signo y el objeto (un ejemplo de esto sería la observación científica de un evento de mimetismo biológico).

Así, la semiótica de Peirce sirve de modelo para la comprensión de determinados fenómenos, desde el comportamiento mimético de otras especies hasta el análisis científico que podemos realizar de esos mismos comportamientos. Al parecer, esta es una de las pocas teorías capaces de abordar estas manifestaciones comunicativas variadas presentes en el mundo natural.

#### **4. Comprensión de algunos ejemplos de mimetismo biológico a través de la semiótica peirceana**

Para el análisis de algunos casos de mimetismo desde la semiótica de Peirce es necesario definir lo siguiente: el representamen [R] debe entenderse como el Imitador o como he preferido llamarlo acto mimético; no es el animal en sí mismo, sino la acción de imitar la que es identificada como el representamen. Luego, el Objeto [O] será el modelo, esto quiere decir cualquier animal que pueda ser imitado por otro y finalmente el



interpretante [I] se entenderá como el Operador. Esta interpretación de los actos miméticos fue construida por Vane-Wright en 1976 (Maran 2017) y Maran la retoma para complementar su teoría. La comprensión de esta relación triádica se esclarece a través de un esquema que presenta las relaciones tradicionales de la semiótica peirceana (véase figura 1). Explicado el contexto teórico de esta propuesta, se revisarán algunos casos de mimetismo a partir de la estructura presentada.

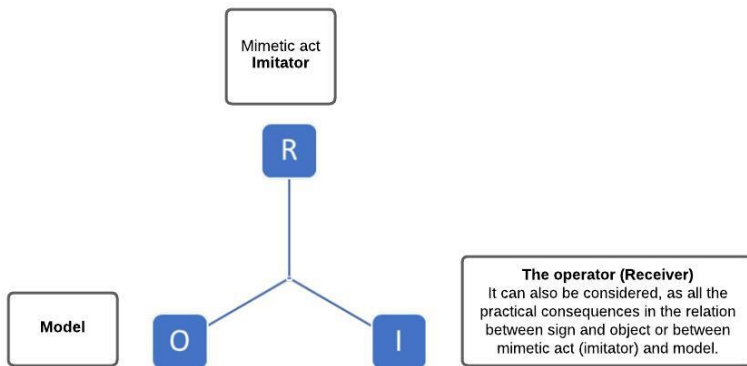


Figura 4: Tripié basado en la relación de la triada semiótica de Peirce junto con la interpretación de los elementos de los actos miméticos construida por Vane-Wright.

### Un caso de mimetismo simbiótico en *Maculinea arion* (mimetismo agresivo)

El tipo de mimetismo que se analizará a continuación es común a las especies parásitas que suelen imitar una apariencia inofensiva para obtener un beneficio particular de su huésped. El ejemplo de la mariposa azul europea *Maculinea arion* es ilustrativo. La hembra deja sus huevos en la planta de tomillo, después de eclosionar caen al suelo en forma de larvas e imitan a través de compuestos volátiles el olor de la larva de la especie de hormiga *Myrmica sabuleti*. Estas hormigas llevan las larvas de mariposa al hormiguero donde procederán a alimentarlas como si fueran una larva de hormiga más e incluso en su avanzado desarrollo se alimentarán incluso de hormigas jóvenes. La metamorfosis se produ-

ce dentro de la colonia de hormigas de la que emerge como mariposa. La conservación de este tipo de mariposa en Gran Bretaña fue posible al comprender el *modus operandi* de su reproducción y metamorfosis parasitaria y a la vez simbiótica ya que esta mariposa no causa daños significativos que comprometan la colonia de hormigas (Thomas 1995; Gilbert et al. 2012). En este ejemplo, el signo o representamen (el acto mimético o imitador) se da a través de los compuestos volátiles que copian las sustancias químicas de la larva de hormiga.

Siguiendo el ejemplo anterior, el uso de los componentes de la tríada en la propuesta de Maran se aplicaría de la siguiente manera:

- **Representamen/acto mimético/imitador:** secreción por las larvas de la mariposa *Maculinea arion* de sustancias químicas que utilizan como modelo a las larvas de hormiga.
- **Objeto/Modelo:** larva de hormiga *Myrmica sabuleti*.
- **Interpretante/receptor:** depredadores potenciales que como consecuencia práctica identifican mariposas distintas a las mencionadas anteriormente para alimentarse.

El traspaso de información favorece a la mariposa *Maculinea arion*, generando un hábito de acción que se relaciona directamente con la prosecución del ciclo vital de la mariposa. Este tipo de relaciones es común en los organismos vivos. La interacción comunicativa favorece medios de supervivencia.

### **Mimetismo mülleriano: el caso de la mariposa *Heliconius numata***

Otro tipo de mimetismo ocurre en la mariposa *Heliconius numata* que presenta siete patrones diferentes de coloración de alas, uno de los cuales imita el de la mariposa *Melinaea mneme*, una especie relacionada. Investigaciones del genoma de estas mariposas demostraron que el cambio en el patrón cromático de las alas de la mariposa *Heliconius numata* se debe a un súper gen que a través de la recombinación de sus secciones produce la manifestación de diversos patrones, uno de los cuales imita la coloración de las mariposas *Melinaea mneme* que sabe mal a sus depredadores. Con este mimetismo, el beneficio de no ser depredado se replica para *Heliconius numata* (Joron et al. 2011; Lanteri & Del Río 2014). De este modo, semióticamente se podría analizar como sigue:

- **Representamen/acto mimético/imitador:** la posibilidad genética de la mariposa *Heliconius numata* de mimetizarse con la mariposa *Melinaea mneme*.
- **Objeto/Modelo:** mariposa *Melinaea mneme*.
- **Interpretante/receptor:** posibles depredadores que como consecuencia práctica no comerán a las mariposas que manifiesten a través de sus colores el patrón imitador.

Este ejemplo es paradigmático porque representa aquella comunicación como traspaso de información no intencionada que repercute prácticamente en los hábitos de acción de seres vivos. Lo interesante de este ejemplo es la continuidad causal presente en el fenómeno analizado. La prevalencia en la mariposa *Melinaea mneme* del patrón mimético respalda la idea de una cadena de semiosis donde los depredadores, al inferir la posibilidad de que la mariposa tenga mal sabor, favorecen la mantención de la especie al no ser depredada.

## Dos ejemplos de mimetismo batesiano

El mimetismo batesiano se caracteriza por un animal, insecto o planta que imita la apariencia o el comportamiento de algún modelo que puede resultar peligroso o aversivo para su depredador. Así, una forma de mimetismo interesante para tener en cuenta lo encontramos en la planta *Pulmonaria officinalis*, originaria del suroeste de Inglaterra (Farmer 2014 en Pannell & Farmer 2016), que presenta patrones blanquecinos en sus hojas que imitan los excrementos de las aves y esto la libera de sus depredadores. En este caso, su intérprete (si el mimetismo funciona) debería ser un intérprete inmediato ya que las consecuencias prácticas se derivarán del desconocimiento de la caracterización mimética por parte de los potenciales depredadores. Así, la tríada semiótica se aplicaría de la siguiente manera:

- **Representamen/Acto mimético/imitador:** manchas blanquecinas en *Pulmonaria officinalis*.
- **Objeto/Modelo:** excrementos de pájaros.
- **Interpretante/receptor:** posibles depredadores que evitarán comerse la planta por aversión.

En este ejemplo es interesante mencionar nuevamente lo que Dennett propone sobre el enfoque de diseño. Preguntarse cómo la planta adquiere evolutivamente estos patrones en sus hojas es muy interesante para poder determinar la cadena de semiosis que finalmente favorece que la *Pulmonaria Officinalis* no sea depredada.

Otro ejemplo del mimetismo batesiano es el comportamiento del pulpo *Thaumoctopus mimicus* descubierto frente a las costas de Indonesia en 1998 (Lanteri & Del Río 2014); este imita con su coloración y con el movimiento ondulatorio de sus tentáculos a varias especies de peces, serpientes, anémonas, cangrejos y estrellas de mar que comparten el mismo entorno, todos ellos venenosos en distintos grados. En este caso, lo que más llamó la atención de los investigadores es la gran cantidad de especies modelo para desarrollar comportamientos miméticos.

- **Representamen/Acto mimético/imitador:** imitaciones del pulpo *Thaumoctopus mimicus*.
- **Objeto/Modelo:** varias especies venenosas.
- **Interpretante/receptor:** posibles depredadores.

Este caso es claramente diferente a los otros revisados porque podemos identificar intencionalidad en las acciones miméticas del pulpo al cambiar su imitación volitivamente dependiendo del tipo de depredador al que se enfrente. De este modo, cuando el pulpo infiere las posibilidades de acción al enfrentarse a un depredador, este no es un interpretante inmediato, más bien es un interpretante dinámico por su complejidad. El acto mimético sería no solo icónico sino que también indexical, al indicar un peligro activo al depredador y activando de este modo la cadena de semiosis.

Los casos de mimetismo revisados tienen relevancia en primer lugar para la comprensión de estos fenómenos a través del concepto de ícono, así como para la profundización del análisis que sea posible a partir de los interpretantes que se identifican en los diversos ejemplos revisados. Se podría afirmar que el carácter inseparable y no jerárquico de los elementos que componen la semiosis, por un lado, dificulta la comprensión de esta propuesta por razones relacionadas con el contexto teórico epistémico en el que generalmente apreciamos los fenómenos de la realidad. Por otro lado, la propuesta semiótica pragmatista nos desafía a concebir versiones conceptualmente más complejas de ciertas nociones como: inferencia, comunicación, conducta, intencionalidad, conciencia, entre otras.

También vale la pena señalar que algunos de los ejemplos revisados muestran comportamiento más complejos e intencionales que otros. El caso de la mariposa *Heliconius numata* es único, ya que la diversidad de patrones se debe a la expresión de un súper gen, pero la preponderancia de un patrón particular se debe al cambio de hábito de los depredadores de la mariposa, por lo que el proceso de mimetismo no es mérito de la mariposa, sino de la interacción con sus depredadores. Lo interesante, como se hizo explícito, es que la semiótica peirceana da cuenta del carácter relacional del signo sirviendo como modelo.

## 5. Reflexión final y conclusiones

En biosemiótica (como extensión de la semiótica pragmaticista) el dominio de los posibles interpretantes depende de un mundo circundante (*umwelt*). Los interpretantes cambian sus hábitos de acción según la relación mimética que se establece con otro, sea este un animal, insecto, planta u hongo. Siguiendo a Favreau:

*“[...] the ubiquity of sign processes in nature, as evidenced across a wide spectrum of semiotically interacting [...]”* [la ubicuidad de los procesos de signos en la naturaleza, tal y como se evidencia en un amplio espectro de interacciones semióticas] (2007, p. 67) permite la comprensión de ciertos procesos como parte de la comunicación en un sentido amplio. Así, dentro de esta propuesta, los seres y objetos del mundo como signos son parte del *modus operandi* de la vida para su mantenimiento vital, a través de relaciones que forman una red de comunicación interespecies y retroalimentan la biodiversidad del medio circundante (*umwelt*) desde múltiples cadenas de semiosis. Esto implica la construcción comunitaria de conocimiento en el mundo natural en diferentes niveles, pero con ciertos márgenes definidos por estructuras químicas, biológicas y físicas que a veces interactúan de forma mecánica, pero en otros muchos casos de formas más complejas y no deterministas.

Es posible afirmar que en la semiosis no solo interactúan seres humanos, ya Peirce, Morris y posteriormente Sebeok dieron cuenta de la propuesta en la cual participan animales de otras especies; Peirce se acercó a esta perspectiva como se puede apreciar en el siguiente fragmento:

*“Uncontrolled inference from contiguity or experiential connection is the most rudimentary of all reasoning. The lower animals so reason. A dog when he hears his master’s voice runs expecting to see him; and if he does not find him, will manifest surprise or at any rate perplexity.”* [La inferencia incontrolada a partir de la contigüidad o conexión experiencial es el más rudimentario de todos los razonamientos. Los animales inferiores así lo razonan. Un perro, al oír la voz de su amo, corre esperando verlo; y si no lo encuentra, manifestará sorpresa o en todo caso perplejidad.] (CP 7.445)

Si un ser vivo es capaz de generar inferencias, ¿podría decirse que la capacidad de ser interpretante está incorporada a los procesos comunicativos *per se*? ¿será esto una de las condiciones para todo ser vivo?

Algunas conclusiones de este escrito son las siguientes: en primer lugar, que la estructura semiótica peirceana consolida su máxima expresión como herramienta teórica cuando se utiliza en constante relación con ciertos elementos propios de la epistemología de Peirce, a saber, su máxima pragmática y su sistema categorial.

La semiótica del final de la vida del filósofo es lo suficientemente compleja como para servir de modelo para una amplia gama de procesos comunicativos, comprendiendo desde los componentes emocionales ligados a la comunicación hasta formalizaciones complejas que dan cuenta de abstracciones creadas simbólicamente para la comprensión.

En segundo lugar, los ejemplos presentados exponen cómo la semiótica peirceana se ha constituido como el marco fundamental para la biosemiótica, siendo esta última una extensión de esta al proporcionar elementos teóricos para la comprensión de formas comunicativas o de transferencia de información no humanas. En los ejemplos revisados, el objetivo del proceso de imitación radica en no ser depredado o en un proceso que favorezca el mantenimiento y desarrollo vital de los individuos; esto se da muchas veces a partir de inferencias por parte del operador o interpretante lo que genera cambios de hábitos. El alcance de comprender los procesos de mimetismo biológico a través de la semiótica peirceana tiene implicaciones en la comprensión de ciertas cuestiones, por ejemplo, en cómo el trabajo científico establece relaciones que antes no era posible desarrollar, como el mimetismo biológico utilizado por nuestra especie a través de las vacunas de ARNm para Covid-19 y el desarrollo de una respuesta inmune que nos proteja de los efectos más agudos de la enfermedad. Tipos similares de mecanismos

miméticos ya han sido estudiados por biólogos moleculares, quienes los han considerado estrategias de supervivencia (Ariza & Gómez 2017).

También hay investigaciones en las que se utiliza la tríada semiótica para desarrollar conocimientos biocomunicativos. Como se indicó en el primer ejemplo de la mariposa *Maculinea arion*, la comprensión de algunos procesos comunicativos interespecie puede ayudar a decidir acciones hacia la preservación de especies que utilizan actos miméticos como forma cooperativa de subsistencia (Thomas 1995).

Poner de relieve nuevas perspectivas sobre la comunicación es uno de los objetivos principales de este tipo de investigaciones. La noche eterna del sujeto cartesiano debe abrir paso a una transubjetividad que dé cuenta de la sempiterna relación de cada ser vivo con su entorno. El modelo biosemiótico nos ayuda a comprender la pluralidad y complejidad de las relaciones interespecie, sean estas intencionadas o no. No estamos solos en el pensamiento; por el contrario, somos en la semiosis, sin ésta la existencia simplemente sería inverosímil.

## Referencias

Anderson, M., Deely, J., Krampen, M., Ransdell, J., Sebeok, T., & von Uexküll, T. (1984). A semiotic perspective on the sciences: Steps toward a new paradigm. *Semiotica*, 52(1-2).

Ariza-Mateos, A., & Gómez, J. (2017). Viral tRNA Mimicry from a Biocommunicative Perspective. *Front. Microbiol*, 8 , pp. 1-14. doi: 10.3389/fmicb.2017.02395

Atkin, A. (2005). Peirce On The Index and Indexical Reference. *Transactions of The Charles S. Peirce Society*, 41(1), pp. 161–188.

Atkin, A. (2022). Peirce's Theory of Signs. *The Stanford Encyclopedia of Philosophy*, (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.) <https://plato.stanford.edu/entries/peirce-semiotics/>

Barbieri, M. (2009). A short history of biosemiotics. *Biosemiotics*, 2, pp. 221-245.

Barbieri, M. (2014). From Biosemiotics to Code Biology. *Biological Theory*, 9(2), pp. 239–249. doi:10.1007/s13752-013-0155-6

Barrena, S., & Nubiola, J. (2013). Charles S. Peirce (1839- 1814): Un pensador para el siglo XXI. Editorial Eunsa.

Deely, J. (2001). Umwelt. *Semiotica*, 134, pp. 125-135. doi:10.1515/semi.2001.019

Dennett, D. (2024). De las bacterias a Bach. La evolución de la mente. Ediciones de Pasado y Presente.

Favreau, D. (2007). Fundaments of Animal Knowing: Establishing Relations Between Sensations, Actions, and the World in Guenther Witzany (Ed.) *Biosemiotics in Transdisciplinary Contexts* (pp. 61-69). Umweb Press.

Gianoli, E., & Carrasco-Urra, F. (2014). Leaf Mimicry in a Climbing Plant Protects against Herbivory. *Current Biology*, 24(9), pp. 984–987. doi:10.1016/j.cub.2014.03.010

Gilbert, S., Sapp, J., & Taubert, A. (2012). A symbiotic view of life: we have never been individuals. *The quarterly review of biology*, 87(4), pp. 325–341.

Godfrey-Smith (2014). *Philosophy of Biology*. Princeton University Press.

Houser, N. (2014). The Intelligible Universe in V. Romanini & E. Fernández (Ed.), *Peirce and Biosemiotics A Guess at the Riddle of Life* (pp.14-45). Springer

Joron, M., Frezal, L., Jones, R., et al. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, 477, pp. 203–206. <https://doi.org/10.1038/nature10341>

Kull, K., Deacon, T., Emmeche, C., et al. (2009). Theses on Biosemiotics: Prolegomena to a Theoretical Biology. *Biology Theory*, 4, pp. 167–173. <https://doi.org/10.1162/biot.2009.4.2.167>

Kull, K., Deacon, T., Emmeche, C., Hoffmeyer, J., & Stjernfelt, F. (2011). Theses on Biosemiotics: Prolegomena to a Theoretical Biology in K. Kull & C. Emmeche (Eds.), *Towards a semiotic Biology. Life is the Action of Signs* (pp. 25-41). Imperial College Press.

Lalor, B. J. (1997). The classification of Peirce's interpretants. *Semiotica*, 114(1-2). doi:10.1515/semi.1997.114.1-2.31

Lanteri, A., & Del Río, M. G. (2014). La imitación en la naturaleza. *Revista Ciencia Hoy*, 23(138), pp. 56-60.



- Maran, T. (2017). *Mimicry and Meaning: Structure and Semiotics of Biological Mimicry*. Springer.
- Merrell, F. (1997). *Peirce, Signs, and Meaning*. University of Toronto Press.
- Morris, C. (1971). *Writings on the General Theory of Signs*. Mouton & Co. N.V. Publishers The Hague.
- Pannell, J. R., & Farmer, E. E. (2016). Mimicry in plants. *Current Biology*, 26(17), pp. 784–785. doi:10.1016/j.cub.2016.04.005
- Peirce, C. (1992-98). *The Essential Peirce. Selected Philosophical Writings*, vols. 1-2, N. Houser et al. (eds.). Indiana University. EP
- Peirce, C. (1931–1958). *Collected Papers of C.S. Peirce*, 8 vols., ed. by C. Hartshorne, P. Weiss (vols. 1-6) and A. Burks (vols. 7-8). Harvard University Press. CP
- Petrilli, S. (2023). Significs, Pragmatism, and Mother-Sense: Welby's Conversations with Peirce and Others. *European Journal of Pragmatism and American Philosophy*, XV(1) , pp. 1-6. doi: <https://doi.org/10.4000/ejppap.3221>
- Petrilli, S., & Ponzio, A. (2005). *Semiotics Unbounded: Interpretive Routes through the Open Network of Signs*. University of Toronto Press. doi: 10.3138/9781442657113
- Romanini, V. (2014). Semeiosis as a Living Process in V. Romanini & E. Fernández (Eds.), *Peirce and Biosemiotics A Guess at the Riddle of Life* (pp. 215-239). Springer
- Saussure, F. (2011). *Course of General Linguistics*. Columbia University Press.
- Sebeok, T. (1996). *Signos: una Introducción a la Semiótica*. Paidós.
- Short, T. L. (2007). *Peirce's Theory of Signs*. Cambridge UP.
- Thomas, J. A. (1995). The ecology and conservation of *Maculinea arion* and other European species of large blue butterfly in Pullin, A.S. (Eds.), *Ecology and Conservation of Butterflies* (pp. 180-197). Springer. [https://doi.org/10.1007/978-94-011-1282-6\\_13](https://doi.org/10.1007/978-94-011-1282-6_13)
- Uexküll, J. V. (2010). *A foray into the worlds of animals and humans*. University of Minnesota Press.



# Algunas observaciones críticas a la noción de modelo en el Enfoque Heredado

Juan Redmond<sup>a</sup>

## Resumen

En el presente artículo presentamos algunas observaciones críticas a la noción de modelo en el llamado enfoque heredado de la filosofía de la ciencia.

**Palabras clave:** Modelo, Enfoque Heredado, representación, lógica.

## Abstract

In this paper we present some critical remarks on the notion of model in the so-called received view in the philosophy of science.

**Keywords:** Model, Inherited Approach, representation, logic.

En los libros de literatura clásica sobre filosofía de la ciencia se indica que, con el desarrollo del llamado empirismo lógico (círculo de Viena), surge una noción original de teoría científica. El enfoque que emerge del trabajo de estos autores es conocido en nuestros días como el Enfoque Heredado. En el núcleo de una teoría científica, de acuerdo con estos autores, se ubican axiomas y principios generales expresados en el lenguaje formal de la lógica y acompañados por una relación de consecuencia lógica que permite derivaciones deductivas a partir de ellos<sup>1</sup>. En este sentido se dice que una teoría científica  $[T]$ , en términos

---

<sup>a</sup>Universidad de Valparaíso, Valparaíso, Chile.

Contacto: [juan.redmond@uv.cl](mailto:juan.redmond@uv.cl)

<sup>1</sup>En nuestros días esta idea estaría puesta en cuestión por núcleos de conocimiento donde no pueden reconocerse ni leyes ni axiomas. Sería el caso, por ejemplo, de la Biología.

formales, es un conjunto deductivamente clausurado  $[D]$  de afirmaciones:  $D(T)$ . Es decir,  $D(T)$  contiene todas las deducciones que pueden hacerse a partir de  $T$ . Así, una teoría  $T$  se identifica con la clausura deductiva de sus axiomas. Siguiendo a Carnap (1938, p. 199), cuando estos axiomas son interpretados aparece otro elemento fundamental de una teoría científica: las leyes. A su vez, el lenguaje propio de la teoría, expresado lógicamente, se divide en términos lógicos (conectivas, cuantificadores) y no lógicos. Estos últimos se dividen en observacionales: términos que se interpretan como refiriéndose a objetos observables; y no observacionales o términos teóricos (energía, protón, etc.), que se interpretan como conectados a la observación por medio de reglas de correspondencia.

Interpretar un lenguaje formal significa construir una relación entre los elementos de ese lenguaje y algo externo al lenguaje. El Enfoque Heredado sostiene, siguiendo los lineamientos generales del pensamiento lógico de Tarski, que el lenguaje formal de una teoría se interpreta por medio de modelos<sup>2</sup>. Analicemos a continuación el rol que desempeñan los modelos en la construcción del significado de una teoría según el enfoque heredado.<sup>3</sup>

## 1. Modelos formales y Modelos representacionales

Del universo de modelos posibles hay dos clases que son bien identificables: los modelos en las ciencias formales, especialmente en lógica y matemática, y los modelos que emergen de la práctica científica que se propone representar porciones de los fenómenos. Los primeros se denominan modelos formales (lógicos o matemáticos) mientras que los últimos pueden denominarse modelos representacionales (Cf. Frigg 2022, p. 47).

---

<sup>2</sup>“A possible realization in which all valid sentences of a theory  $T$  are satisfied is called a model of  $T$ ” (Tarski 1953, p. 11); “The type of reasoning which has just been applied is known as the METHOD OF PROOF BY EXHIBITING A MODEL” (Tarski 1994, p. 116).

<sup>3</sup>Dejaremos para una futura publicación un análisis crítico de la relación entre los términos teóricos y observacionales y los modelos.

## Modelo formal

Un modelo formal o lógico es una colección de objetos que se encuentran en cierta relación entre sí y a los cuales se les puede atribuir ciertas propiedades. La función u objetivo de una colección de este tipo es interpretar un cierto lenguaje. Interpretar un lenguaje significa poner en relación los símbolos de un lenguaje con los miembros de esta colección. Si un modelo interpreta, por ejemplo, las oraciones de un cierto lenguaje, decimos entonces que estas oraciones son verdaderas. En este sentido se dice que un modelo ofrece una semántica a un cierto lenguaje. Por ejemplo, consideremos la oración  $Pa \Rightarrow Qa$ . La sintaxis que rige este lenguaje nos dice que las minúsculas se refieren a objetos y las mayúsculas a propiedades.  $Pa$  significa que el objeto señalado o referido por  $a$  posee las propiedades  $P$  y  $Q$ , respectivamente. Estas indicaciones de la sintaxis son puramente formales. Dar una interpretación a esta oración significa hallar una colección que la haga verdadera. Por ejemplo, si consideramos un modelo donde  $a =$  Gabriela Mistral,  $P =$  nacida en Vicuña, provincia de Elqui y  $Q =$  nacida en Chile, diremos que la interpretación de  $Pa \Rightarrow Qa$  en este modelo es verdadera, pues es verdadero que ‘si Gabriela Mistral nació en Vicuña, entonces nació en Chile’.

En principio lo que podemos inferir de este modo de entender la interpretación de un lenguaje formal por medio de modelos, es que puede haber tantas como uno quiera. Y esta variedad puede ser tan diversa como uno quiera. Pero lo notable aquí, desde nuestro punto de vista, es que, si bien el propósito de estos modelos no es representacional, están interpretando axiomas y leyes que sí quieren dar cuenta de la regularidad de los fenómenos. Más abajo nos referiremos a esto último.

## Modelo representacional

Un modelo representacional es aquel que está destinado a representar algo. No interpreta un lenguaje, sino que su función es la de subrogar una porción fenoménica. Cómo sea esta subrogación depende del enfoque de representación científica que consideremos (Cf. Frigg & Nguyen 2017). Pero ninguna de estas funciones es la de interpretar un lenguaje. Aquello que representa el modelo se denomina el sistema objetivo del modelo (target system) y que normalmente constituye una porción fenoménica. De entre todas las porciones fenoménicas distinguimos en nuestro trabajo un subconjunto que llamamos porciones problemáticas de los fenómenos y las consideramos del mayor interés

de estudio porque son las que presentan, desde nuestro punto de vista, los mayores desafíos filosóficos sobre la práctica de modelización.

En general, creemos que hablamos de porciones problemáticas de los fenómenos en los siguientes casos<sup>4</sup>:

1. sabemos que hay algo, pero no sabemos realmente cómo es [Si pensamos, por ejemplo, en las diferentes formas dadas al átomo desde Dalton hasta Bohr];
2. no sabemos si hay algo o no [si pensamos en el Modelo Planetario que proponía el planeta Vulcano entre Mercurio y el Sol];
3. sabemos que tal cosa no existe [modelos de agentes perfectos (los agentes perfectos no existen)].

De acuerdo con Frigg y Nguyen (2017), toda perspectiva sobre la representación científica debe ser capaz, en primer lugar, de rellenar el espacio en blanco del siguiente esquema:

‘ $R$  es una representación científica del sistema objetivo  $SO$   
si y solo si \_\_\_\_.’<sup>5</sup>

Esta fórmula es conocida como el Problema de la Representación Científica/Epistémica (Scientific Representation Problem; Epistemic Representation Problem). La diferencia entre científico y epistémico se refiere a lo que Callender y Cohen (2006, pp. 68-69) señalan como el “problema de demarcación” (siguiendo a Popper) para las “representaciones científicas”, para quienes demarcan las representaciones científicas de las que no lo son; y “representaciones epistémicas” para quienes consideran irrelevante esa distinción, siguiendo la sugerencia de Contessa (2007) de ampliar el alcance de la investigación.

Otros temas y condiciones a los cuales debe responder una práctica de modelización para que sus modelos sean considerados como representacionales son las siguientes:

---

<sup>4</sup>Cf. Redmond 2021, p. 302

<sup>5</sup>Hemos reemplazado aquí ‘ $S$ ’ (scientific representation) por ‘ $R$ ’, y también ‘ $T$ ’ (target system) por ‘ $SO$ ’ (sistema objetivo), del esquema original de Frigg y Nguyen (2017).

- El problema de la demarcación representacional: la cuestión de cómo las representaciones científicas difieren de otros tipos de representaciones.
- El problema del estilo: ¿qué estilos hay y cómo se pueden caracterizar?
- La formulación de normas de exactitud: ¿cómo identificar lo que constituye una representación exacta?
- El problema de la ontología: ¿qué tipo de objetos sirven como representaciones?

Y de modo adicional, como condiciones de adecuación o de suficiencia se deben tener en consideración las siguientes cuestiones:

- Razonamiento sustituto: los modelos como representaciones científicas deben permitir generar hipótesis sobre sus sistemas objetivo.
- Posibilidad de distorsión (misrepresentation): si  $R$  no representa exactamente a  $SO$ , entonces se trata de una distorsión pero no de una no-representación.
- Modelos sin objetivo (Targetless): dar respuesta a la pregunta ¿qué debemos hacer con las representaciones científicas que carecen de targets, pero que son exitosos de igual manera?
- Requisito de direccionalidad: las representaciones científicas se refieren a sus targets, pero los targets no se refieren a sus representaciones.
- Aplicabilidad de las matemáticas: ¿cómo el aparato matemático utilizado en algunas representaciones científicas se vincula o enlaza con el mundo físico?

Destacaremos a continuación los tres enfoques sobre la representación científica que, desde nuestro punto de vista, son relevantes para nuestro análisis: el enfoque del Semanticismo (Suppes 1960, 1970; Stegmüller 1970; Balzer et al. 1987; Suppe 1989), el Enfoque Similarista de Ronald Giere (año) y el Inferencialismo de Mauricio Suárez (2004). De acuerdo con el Semanticismo, la noción de representación se enmarca en la práctica de reconstrucción teórica y matemática de las teorías científicas en términos de una clase o conjunto de modelos definidos como representaciones adecuadas de los fenómenos (Díez

& Moulines 2016, p. 348). Según este enfoque, estas representaciones, cuando se aplican efectivamente al mundo, permiten a las teorías describir las características de los fenómenos del mundo. Como demostración, los semanticistas construyen caracterizaciones axiomáticas bastante eficientes (en términos de teoría de conjuntos) de la relación entre un modelo y el sistema de fenómenos. Estas últimas son definidas en términos de un morfismo (homomorfismo, isomorfismo, entre otros). Esto supone la existencia de estructuras matemáticas y fenoménicas, con las que se ponen en relación. Es decir, esta relación se entiende como una relación diádica de correspondencia (o identidad, aproximación, subsunción, etc.) entre las estructuras representativas y las estructuras fenoménicas a las que se dirigen. Tratándose de modelos representacionales y no modelos formales, se trataría de encontrar las mejores caracterizaciones de cómo es que un modelo científico  $M$  representa a su objetivo  $TS$ . Pero esto último ocurre si y sólo si  $TS$  es isomorfo a  $M$ .

En el caso del enfoque similarista (o perspectiva cognitivista) de Giere, si bien se inscribe en la discusión estructuralista sobre la representación, ahora hace hincapié en los diversos usos que los científicos hacen del término, ya no en términos lógico-matemáticos. Para Giere, los modelos, más que estructuras o entidades matemáticas (entidades conjuntivas), son entidades de naturaleza más amplia y sin forma determinada (diagramas, dibujos, mapas, organismos, etc.), que se utilizan a partir de una conexión diferente a la de las conexiones matemáticas o conexiones lógicas (en términos conjuntivos): la idea de similaridad. La noción estructuralista de isomorfismo entre los modelos y el mundo pretende captar sólo una de las formas en que los científicos utilizan los modelos. Los modelos sólo serían similares en ciertos aspectos y en grados suficientes en función de su uso y según un contexto epistémico específico (Giere 1988, p. 81). Asimismo, las relaciones que se establecen entre un modelo y el sistema real al que se dirigen también serían relaciones de semejanza: “Un sistema real es identificado como semejante a uno de los modelos [de la teoría]” (Giere 1988, p. 86). Giere propone que la función representacional de un modelo, entendida como similaridad, lleva a pensar los modelos en términos cognitivos y pragmáticos: “son ‘mapas internos’ del mundo externo” (Giere 1988, p. 6). Esto implicaría más bien tomar en consideración los factores cognitivos que intervienen en las estrategias de los científicos para construir mapas del mundo con el fin de cartografiar, y así resolver, los problemas que surgen de los fenómenos que estudian. En conjunto, Giere no intenta ofrecer una de-



finición precisa de la similaridad: el éxito de nuestras representaciones del mundo se sustenta en nuestras capacidades cognitivas (lenguaje, atención, percepción, imaginación, explicación, etc.). Así pues, para estudiar el éxito de una representación, hay que estudiar los resultados de las ciencias cognitivas, lo que permite demostrar y explicar el éxito del esfuerzo del conocimiento científico a gran escala. Su enfoque cognitivista de la ciencia muestra así que el problema de la representación está relacionado con las capacidades naturales de los agentes del conocimiento en el contexto de la utilización de los recursos disponibles para representar el mundo. Por lo tanto, el agente de conocimiento seguiría estando en el centro de una teoría cognitiva de la ciencia, término que utiliza para referirse a su enfoque. “Que los humanos (y los animales) crean representaciones internas de su entorno (así como de sí mismos) es probablemente la noción central en las ciencias cognitivas” (Giere 1988, p. 6). Sin embargo, esta noción de representación para los modelos que acuña Giere no parece estar alineada con las exigencias del Enfoque Heredado. En efecto, en esta última no es posible pensar los modelos, sino que conteniendo una estructura (ver más abajo) que guarda una relación isomorfa con aquello que representa. Más distante aún se encuentra la perspectiva inferencialista de Suárez.

Con el inferencialismo de Mauricio Suárez (2004), la noción de representación se desplaza, y el foco se vuelve hacia las inferencias sustitutivas sobre fenómenos a partir de modelos. Los modelos se definen entonces como herramientas que nos permiten apuntar y generar hipótesis plausibles sobre sus sistemas de fenómenos. Esta idea sugiere una definición deflacionaria de la representación. Así, la función primordial de un modelo es su función inferencial, que crea la posibilidad de adquirir conocimientos sin examinar directamente el  $ST$ , sino mirando directamente a  $M$ , siempre que  $M$  sea “coherente.” “aborde.”<sup>el</sup> sistema objetivo en los aspectos apropiados y en grado suficiente. La afirmación anterior supone que las inferencias tienden puentes entre el “mundo modeloz los fenómenos, ya sean deductivas, inductivas o abductivas (López Orellana 2020). La modelación científica procede de esta manera y depende sólo de la habilidad -o agencia- del científico para señalar y hacer inferencias a partir de modelos. La representación científica se define entonces de la siguiente manera:  $M$  representa  $TS$  sólo si la fuerza representacional de  $M$  apunta hacia  $TS$ , y  $M$  permite a agentes competentes e informados hacer inferencias específicas respecto a  $TS$  (Suárez 2004, p. 773).

Tenemos entonces que, en términos generales, el proceso de modelización se puede esquematizar del siguiente modo:

1. El modelo es presentado e identificado por medio de descripciones cuantitativas (matemática) y/o cualitativas (propiedades, relaciones, funciones, etc.), ambas recuperadas tanto de la información que aporta la porción problemática de los fenómenos que es objeto de nuestro estudio (mayormente por la vía de las mediciones) como de diferentes enfoques teóricos (Física, Química, Biología, etc.).
2. A partir de los datos del modelo se infieren conclusiones según distintos tipos de ‘relaciones inferenciales’ (deducción, inducción, abducción, etc.).
3. Lo inferido en el modelo es transpuesto y cotejado o evaluado –en carácter de hipótesis– en la porción problemática de los fenómenos.

## 2. Observaciones críticas

Lo que resulta destacable, desde nuestro punto de vista, es la relación que guardan estos dos tipos de modelos entre sí. En principio son independientes el uno del otro, es decir, algo puede cumplir la función de modelo formal sin ser representacional y viceversa. Pero, sin embargo, un modelo puede cumplir con las dos funciones al mismo tiempo, como ya señalaba Hesse (1967, p. 354). En efecto, como señalamos más arriba, un modelo formal se propone interpretar un lenguaje formal, pero no es excluyente que al mismo tiempo cumpla con una función representacional.

Ahora bien, nos preguntamos, ¿qué sentido tendría dar un significado por medio de modelos si estos no representan a los fenómenos? La respuesta que dieron los gestores del enfoque heredado fue clara: el uso de modelos brindaría la posibilidad de dar significado al formalismo de una teoría con nociones y objetos familiares y que facilitarían el entendimiento. Es el caso, entre otros, de Nagel y Braithwaite que recalcan las dotes heurísticas de los modelos para la construcción de teorías (Cf. Nagel 1961; Braithwaite 1953). Pero al mismo tiempo ambos advirtieron sobre los peligros de dejarse llevar por las representaciones asociadas al uso de estos modelos. Nagel los califica de “potential intellectual trap” (1961, p. 115). Lo que subyace a estas consideraciones

es que estos modelos no están destinados a mediar entre los axiomas de la teoría y el mundo del que habla la teoría. En palabras de Frigg:

Models as construed in the Received View are not used representationally, and they play no role in bringing about, or even understanding, the theory's relation to the world. The theory relates to its subject matter through observation terms and correspondence rules, and models are immaterial in this. (Frigg 2022, p. 51)

Frigg señala que diferentes autores consideraron en su momento que esta noción de modelo del empirismo lógico, no hacía justicia a los casos de modelos que se encuentran en las ciencias (especialmente en la Física). A los defensores del enfoque heredado –señala Frigg (2022, p. 54)– no les habría impresionado esta crítica porque consideran que ni el análisis de lo que piensan los físicos ni el análisis de los ejemplos de la práctica científica son competencia de la filosofía de la ciencia. En efecto, la prioridad para estos autores, al parecer, la tiene el lenguaje formal que representa los axiomas y las leyes que dan cuenta de la regularidad de los fenómenos. Lo que queda claro para estos autores es que los modelos no subrogan a porciones fenoménicas y por ende no hay relación inferencial entre ellos. Es decir, no hay razonamiento subrogado a partir del modelo y sobre su sistema objetivo (target system). Los encadenamientos lógicos que llevan a conclusiones sobre los fenómenos se realizarían desde el sistema axiomático de base.

Autores que se opusieron a esto último fueron Achinstein, Spector y Swanson<sup>6</sup> que sostenían que los modelos de las teorías debían ser representacionales. De este modo los modelos eran pensados como llevando una relación sustantiva con el mundo que pretendían representar. Pero incluso en estos autores, de acuerdo a nuestro punto de vista, no estaría permitido hacer lógica desde el modelo. Es decir, si bien los modelos se corresponden representacionalmente con los fenómenos, no sería desde ellos desde donde justificamos lógicamente nuestras afirmaciones sobre los fenómenos sino también desde el sistema axiomático de base.

---

<sup>6</sup>Cf. Frigg 2022, p. 54

### 3. Modelos y estructuras

Las nociones de modelo formal y estructura van de la mano. Una estructura se define como entidad compuesta formada por (i) un conjunto no vacío  $U$  de objetos denominado dominio (o universo) de la estructura y (ii) un conjunto indexado  $R$  (es decir, una lista ordenada) de relaciones sobre  $U$ .

Por ejemplo, definimos la estructura  $E_1 = (U, R)$  como compuesta de un dominio  $U$  donde están las entidades  $a_1$  y  $a_2$ , y las relaciones  $R_1$  (binaria) que extensionalmente la definimos como  $(a_1, a_2)$ , y  $R_2$  (monádica) que la definimos como  $\{a_1\}$ .

Definida de este modo se denomina ‘estructura teórico-conjuntista  $S$ ’ (set-theoretical structure). Ciertamente podemos agregar otros elementos a esta base estructural  $S$  y obtener otro tipo de estructuras. Frigg (2022, p. 56) señala dos variantes:

1. Agregando a  $S$  operaciones que mapean elementos del dominio sobre ellos mismos. Al igual que las relaciones, se presentan extensionalmente.
2. Agregando una colección de símbolos (la signatura  $\Sigma$ ) más una función interpretación.

Las estructuras del tipo 2 se llaman modelo-teoréticas. Estas estructuras contienen símbolos denotando partes de la estructura, pero no tiene elementos sintácticos para expresar afirmaciones o un argumento.

Un ejemplo de esta última sería el siguiente: dada la estructura conjuntista del caso anterior:  $E_1 = (U, R)$ , le agregamos la signatura  $\Sigma = \{\sigma_{a1}, \sigma_{a2}, \sigma_{R1}, \sigma_{R2}\}$  y la interpretación  $I$  que dice que  $\sigma_{a1}$  se refiere a  $a_1$ ,  $\sigma_{a2}$  se refiere a  $a_2$ ,  $\sigma_{R1}$  se refiere a  $R_1$ , etc.

De todas estas estructuras las relevantes para el Enfoque Heredado son las teórico-conjuntistas. En efecto, consideremos la afirmación siguiente:  $Fa$  y veamos qué valuación nos da en el Modelo  $E_1$ . Para ello consideremos que  $a$  se refiere a  $a_1$  y  $F$  se refiere a  $R_2$ . De acuerdo con esta Interpretación,  $Fa$  es verdadera pues  $R_2 = \{a_1\}$ .

Con estas herramientas en la mano, ahora se puede estudiar la clase de modelos de una teoría. Consideremos un conjunto de oraciones formales como las presentamos más arriba y consideremos este conjunto como una teoría y lo denotamos por  $T$ . Podemos entonces definir la clase  $C_T$  de todas las estructuras que son modelos de  $T$ . Es decir,  $C_T$

contiene todas las estructuras que hacen que todas las oraciones de  $T$  sean verdaderas bajo alguna interpretación de los términos de  $T$ .

Ahora bien: ¿cómo es esta clase? ¿Qué tipo de estructuras encontramos en esta clase y cómo se relacionan entre sí?

Lo primero que podemos indicar, como señalamos más arriba, es que un conjunto de oraciones formales  $[T]$  puede interpretarse con un número infinito de modelos. De las relaciones que puedan existir entre esos modelos hay dos que resultan relevantes para nuestro estudio: o son isomorfos o no. Si es un conjunto de modelos isomorfos, una Teoría conformada de ese modo tendría un solo modelo y se llamaría Teoría categórica. Otras teorías ganan su significado a partir de modelos, por ejemplo, con diferente número de elementos, es decir, conjuntos de modelos que no son isomorfos, es decir, se trataría de Teorías no categóricas.

#### 4. Desde el modelo a la interpretación

Otro punto importante a considerar es que la práctica científica no parece orientarse siempre desde un conjunto de oraciones formales hacia la búsqueda de un modelo sino al revés. En efecto, los modelos de este tipo que estarían al inicio de esta práctica se llamarían modelos dados (intended models) que, en principio, no permitirían satisfacer ninguna teoría y por ello es necesario buscarla. Es decir, dicho al revés, el punto de partida es una estructura de cierto tipo y vamos a buscar el conjunto de axiomas que la describe. Por el contrario, un modelo sin propósito (unintended model) es aquel que no teníamos previsto o dado y que sí satisface nuestro conjunto de oraciones.

Respecto de los modelos dados, vemos que se corresponde con muchos casos en la ciencia. En efecto, podemos pensar, como ejemplo, en el conjunto de números naturales como modelo que tuvo en Peano, Dedekind y Frege los gestores de sus axiomas. Para estos casos cuesta creer que no tomaran estos modelos como representacionales. De otro modo sería ridículo pensar en que los científicos buscaran sistemas axiomáticos, por ejemplo, para bolas de billar golpeándose sobre un paño (en alusión a la teoría cinética de gases).

Ahora bien, cabe preguntarse, si esta práctica científica a partir de modelos dados es la misma o no que la práctica de modelización que consideramos en nuestros días. En general parecen compartir algunos puntos, pero veamos más en detalle.

1. En la práctica de modelización el punto de partida es un modelo que intenta representar una porción fenoménica. Para el Enfoque Heredado el punto de partida es un modelo dado (intended model) pero no está claro si éste tiene por objetivo representar una porción fenoménica. En el caso de los números naturales, este conjunto considerado como modelo no tiene por objetivo representar nada. Podríamos decir que él mismo constituye una parte del mundo.
2. En la práctica de modelización el modelo es conformado a partir de información tomada de dos lugares: (i) la porción fenoménica que intenta representar, (ii) información teórica que a veces proviene de más de una teoría. En el Enfoque Heredado el modelo permite conformar un sistema axiomático que lo tiene como interpretación y que define una teoría.
3. En la práctica de modelización el modelo es descrito con lenguaje lógico y matemático que permite sacar conclusiones sobre aquel o aquella a quien subroga el modelo (razonamiento sustituto). En el Enfoque Heredado el modelo dado dispara la búsqueda de axiomas cuya interpretación sería este modelo dado. Pero cuán científica sería esta última tarea si no fuera que ese modelo guarda una relación estrecha con los fenómenos. Una explicación posible es que el Enfoque Heredado considera aquí como modelos a porciones representativas de los restantes fenómenos. Al modo como podríamos tomar a nuestro sistema solar como modelo de todos los sistemas planetarios del universo, y buscar las leyes y principios que rigen sus movimientos (Kepler).
4. En la práctica de modelización se establece una relación científica entre el modelo y su sistema objetivo. En el Enfoque Heredado el modelo dado establece una relación con los axiomas buscados.

Pero este sistema axiomático ¿de quién habla? ¿De los fenómenos o solo del modelo que guarda relación con los fenómenos? Una vez más lo repetimos, en este último caso no tendría sentido pensar la tarea a partir de modelos dados si no fueran modelos representacionales. Aunque como ya señalamos más arriba, la tarea emprendida a partir de los Números Naturales no parece ser el caso.

Ahora bien, dado lo anterior, podemos considerar el caso en que en el inicio de ambas tareas se encuentre un modelo. Por ejemplo, para la

práctica de modelización, la propuesta de que el sistema solar es heliocéntrico para resolver problemas de posición de los astros (por ej. El movimiento de retroceso de Marte). Para un/a científico/a que sigue los lineamientos del Enfoque Heredado, se podría tratar de hallar las definiciones, axiomas, leyes, etc., para un modelo dado que no proviene, como en el primer caso, de la búsqueda de explicaciones para una porción fenoménica sino de la búsqueda de una teoría dentro de la cual este modelo dado es una interpretación válida. ¿hasta qué punto estas tareas coinciden?

En principio tenemos la idea de que ambas apuntan a una ampliación de nuestros conocimientos. Pero la calidad de los mismos parece diferente. Podríamos aventurar que mientras que en la primera no sabemos realmente cómo son las cosas y modelamos para saberlo, en la segunda ya sabemos cómo son y buscamos las fórmulas que la describan. Casi diríamos de que se trata de un tema ontológico o de cuáles son los compromisos ontológicos de cada práctica. De acuerdo con la práctica de modelización, no modelamos para probar que algo existe. En todo caso, si el modelo es exitoso podemos creer que hay razones para que la porción fenoménica tenga la forma que anuncia el modelo. Pero esto último, o bien es falso en muchas ocasiones científicas, o bien es imposible de determinar. Mientras que en el Enfoque Heredado diríamos que este compromiso es mayor.

## Referencias

- Balzer, W., Moulines, C. U., & Sneed, J. D. (1987). *An Architectonic for Science. The Structuralist Program*. Synthese Library, Vol. 186. Springer Netherlands.
- Braithwaite, R. B. (1953). *Scientific Explanation*. Cambridge: Cambridge University Press.
- Callender, C, & Cohen, J. (2006). There Is No Special Problem About Scientific Representation. *THEORIA. An International Journal for Theory, History and Foundations of Science*, 21(1), pp. 67-85. <https://doi.org/10.1387/theoria.554>
- Carnap, R. (1938). *Foundations of Logic and Mathematics*. In O. Neurath, C. Morris, and R. Carnap (eds.), *International Encyclopaedia of Unified Science* (Vol. 1). Chicago: University of Chicago Press, pp. 139-213.

Contessa, G. (2007). Scientific representation, interpretation, and surrogate reasoning. *Philosophy of Science*, 74(1) pp. 48-68. <https://doi.org/10.1086/519478>

Díez, J. A., & Moulines, C. U. (2016). *Fundamentos de filosofía de la ciencia*. Editorial Ariel.

Frigg, R. (2022). *Models and Theories: A Philosophical Enquiry*.

Frigg, R., & Nguyen, J. (2017). Models and representation. In L. Magnani & T. Bertolotti (Eds.), *Handbook of model-based science* (pp. 49-102). Springer.

Giere, R. (1988). *Explaining Science: A Cognitive Approach*. University of Chicago Press.

Hesse, M. B. (1967). Models and Analogy in Science. In P. Edwards (ed.), *Encyclopedia of Philosophy*. New York: Macmillan, pp. 354–359.

López-Orellana, R. (2020). Sobre la modelización y la comprensión científicas. Un enfoque inferencial y dinámico aplicado al modelo evodevo *Polypterus* de la plasticidad fenotípica. PhD Thesis. Universidad de Salamanca.

Nagel, E. (1961). *The Structure of Science*. London: Routledge and Keagan Paul.

Redmond, J. (2021). A free dialogical logic for surrogate reasoning: generation of hypothesis without ontological commitments. *THEORIA. An International Journal for Theory, History and Foundations of Science*, 36(3), pp. 297-320. <https://doi.org/10.1387/theoria.21902>

Stegmüller, W. (1970). *Theorie und Erfahrung* (Vol. 2). Springer-Verlag.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science*, 71(5) , pp. 767-779.

Suppe, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. University of Illinois Press.

Suppes, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthese*, 12(2/3), pp. 287-301.



Suppes, P. (1970). *Set-Theoretical Structures in Science*. Stanford University Press.

Tarski, A. (1994). *Introduction to Logic and to the Methodology of the Deductive Sciences*. Oxford University Press.

Tarski, A. (1953). *Undecidable Theories*. Amsterdam: North-Holland Publishing Company.



# Tiempo relacional: bases para una ontología científica minimalista

Franco Reyes Aguirre<sup>a</sup>

## Resumen

El presente artículo tiene como objetivo defender la concepción relacional del tiempo, argumentando que esta permite establecer una ontología científica más parsimoniosa en comparación con la teoría sustantivista. Para sustentar esta tesis, se desarrolla una ontología mínima mediante un enfoque axiomático, que ofrece los fundamentos para formular una ontología científica en la que el tiempo no es considerado como una entidad. El análisis comienza con la caracterización de las entidades existentes, que constituyen el núcleo del relacionismo, e introduce las nociones de mutabilidad, propiedades y estado. Posteriormente, se examina el concepto de cambio, vinculándolo con la idea de devenir. Finalmente, se exploran los conceptos de causalidad y modalidad, y su papel en el marco de una ontología de tiempo relacional.

**Palabras Clave:** devenir, cambio, propiedades, estado.

## Abstract

Abstract: The aim of this article is to defend the relational conception of time, arguing that it enables the establishment of a more parsimonious scientific ontology compared to substantialist theory. To support this thesis, a minimal ontology is developed through an axiomatic approach, which provides the foundations for formulating a scientific ontology in which time is not considered an entity. The analysis begins with the characterization of existing entities, which form the core of

---

<sup>a</sup>Grupo de Estudios de Filosofía de las Ciencias de la Universidad de Chile, Santiago, Chile.

Contacto: franco.reyes.a@ug.uchile.cl

relationism, and introduces the notions of mutability, properties, and state. Subsequently, the concept of change is examined, linking it to the idea of becoming. Finally, the concepts of causality and modality are explored, along with their role within the framework of a relational ontology of time.

**Keywords:** becoming, change, properties, state.

## 1. Introducción

Es ampliamente reconocido que la metafísica de las ciencias se ocupa de las estructuras y principios fundamentales de la realidad sobre los cuales las disciplinas científicas construyen su comprensión del mundo. En este contexto, la naturaleza ontológica del tiempo continúa siendo un gran enigma. Tradicionalmente, la pregunta por la naturaleza del tiempo ha dado lugar a dos respuestas principales dentro de la metafísica de las ciencias: la teoría relacional y la teoría sustantivalista.

Ciertamente, la física relativista parece respaldar la tesis sustantivalista, ya que en su marco conceptual el espacio-tiempo es tratado como una entidad en sí misma, con propiedades e interacciones con otros entes. Sin embargo, asumir que la teoría de la relatividad zanja el debate es una conclusión apresurada. Como señala Hawley (2009), aunque la relatividad parece inclinarse hacia el sustantivalismo, no establece un consenso claro sobre cuál concepción del tiempo respalda. Por lo tanto, derivar directamente el sustantivalismo a partir de la teoría de la relatividad implica un error metodológico al invertir el orden de justificación.

Con todo, en las últimas décadas han aparecido férreos defensores de la tesis relacional, entre los que destacan autores como Barbour (1982, 1986, 1999), Barbour y Bertotti (1982), Bunge y Maynez (1976), y Rovelli (2018), quienes han articulado sus argumentos con distintos matices y objetivos particulares.

Tal como señaló Pooley (2012, p. 3) *la única consideración fuerte a favor del relacionismo es la navaja de Ockham*. Nuestro motivo para defender el relacionismo va en concordancia con ello. Por consiguiente, este trabajo tiene como objetivo desarrollar una teoría relacional del tiempo que sirva de base para una ontología científica minimalista. Sin embargo, no toda teoría relacional es intrínsecamente más parsimoniosa que cualquier versión del sustantivalismo; por ello, nuestra propuesta no solo se opondrá a la teoría sustantivalista, sino también a cualquier formulación que a nuestro juicio engrose la ontología de manera innecesaria.

En las secciones que siguen procederemos a formular una ontología mínima cuyo objetivo central será demostrar que el concepto de tiempo entendido como una entidad propiamente tal es innecesario. Este enfoque permitirá reducir las estructuras ontológicas a sus componentes esenciales.

## 2. Sobre lo existente: entes y propiedades

Antes de proceder con el desarrollo de esta sección, es pertinente realizar un breve acápite metodológico. En consonancia con los objetivos de este trabajo, he optado por una presentación axiomática, lo que permitirá minimizar la vaguedad y asegurar una conexión clara entre los argumentos. Para ello, emplearemos conceptos fundamentales de la lógica de primer orden, la teoría de conjuntos y la mereología. Aclarado esto, continuemos.

Si pudiésemos a un sustantivista construir su concepción del mundo, es probable que los cimientos de su estructura sean el tiempo y el espacio, o el espacio-tiempo, sobre los cuales pondría el resto de los ladrillos de su ontología. En contraste, los relacionistas, al sostener que el tiempo y el espacio son relaciones entre entidades, debemos situar como punto de partida a las propias cosas. Así, para desarrollar una teoría relacional es fundamental comenzar con una caracterización adecuada de los entes existentes.

Por consiguiente, proponemos el

**Postulado 1: Todo lo que existe es cambiante (o mutable).** Vale decir que todas las cosas (objetos o entes) tienen la propiedad intrínseca de cambiar.

$$\forall x C(x)$$

En esa línea, definiremos el conjunto de todas las cosas existentes tal que:

$$\Theta = \{x : C(x)\}$$

El rasgo distintivo de todas las entidades del mundo es, por tanto, su mutabilidad, concepto que definiremos con mayor precisión más adelante. Por el momento, tomamos como punto de partida la mutabilidad o cambiabilidad de las cosas en el mundo siguiendo la tradición heracliteana del célebre *panta rei* (es decir, todo está en constante cambio).

**Postulado 2: Todas las cosas pueden asociarse mereológicamente para conformar otras cosas.**

Designaremos este proceso de asociación (o suma) mereológica mediante el operador ' $\dagger$ ' con el fin de distinguirlo de la suma aritmética. Así, por ejemplo, la suma mereológica de los objetos  $x$  e  $y$  resulta en la cosa  $z$ , es decir:  $x \dagger y = z$ . Por lo tanto, afirmaremos que  $x$  es parte de  $z$  o, más precisamente:  $x \sqsubset z$ . Del mismo modo, si  $x \dagger y$ , diremos que  $x$  está asociado (o sumado mereológicamente) con  $y$  o bien  $x$  e  $y$  están asociados.

**Postulado 3: Cada cosa posee un conjunto no vacío de propiedades que la dotan de unicidad.**

Con este postulado establecemos que todo ente tiene un conjunto único de propiedades que lo distinguen de otros entes. En consecuencia, proponemos el siguiente principio de individuación:

$$\forall xy \in \Theta(x = y \Leftrightarrow P(x) = P(y))$$

Donde dado el conjunto  $P = \{p_1, \dots, p_n\}$  de todas las propiedades,  $P(x)$  es el subconjunto de  $P$  que posee la cosa  $x$  (o simplemente estado de  $x$ ) y  $P(y)$  el subconjunto de  $P$  que posee  $y$  (o estado de  $y$ ).

**Corolario:**  $\forall x \in \Theta(\exists!P(x))$ , por tanto no existe ninguna cosa idéntica a otra.

**Postulado 4: El universo es la asociación mereológica de todos los entes.**

$$U \equiv \sum_{x \in \Theta} x$$

En coherencia con el Postulado 4, es importante señalar que  $U \neq \Theta$ . La razón de esta distinción radica en que  $\Theta$  en tanto conjunto es un objeto formal que incluye todos los entes existentes sin atribuirles un orden o estructura específicos. En contraste,  $U$  designa a la totalidad del universo concebido como el resultado de la asociación mereológica de todos los entes existentes.  $U$  es un único objeto (compuesto naturalmente), no una colección.

**Corolario 1:** El universo está compuesto de partes.

**Corolario 2:** No existen cosas aisladas, i.e., cada cosa está asociada por lo menos con alguna otra cosa, valga decir:  $\forall xy \in \Theta(x \sqsubset U \rightarrow \exists y(y \dagger x))$ .

**Teorema 1: El universo es una cosa mutable e individualizada por un conjunto único de propiedades.**

**Demostración:** De acuerdo con el Postulado 4, se asume que  $U$  es el resultante de una asociación mereológica. El Postulado 2 establece

que la combinación de entes produce nuevas entidades, mientras que el Postulado 3 afirma que cada cosa posee un conjunto único de propiedades. Finalmente, el Postulado 1 asegura que todas las cosas son mutables. Luego, al combinar estos postulados, concluimos que el universo es una entidad que, en conformidad con la asociación mereológica, es mutable y está individuada por un conjunto único de propiedades.

**Aclaración:** Para evitar paradojas autorreferenciales,  $U$  es considerado un objeto más dentro de nuestro sistema teórico; no obstante, es importante señalar que  $U \notin \Theta$ . Con esto nos aseguramos de no tener problemas de autorreferencia, lo cual no implica que  $U \notin \Theta \rightarrow U$  no existe. Por ende,  $U$  es la excepción al Corolario 2 del Postulado 4.

Un ejemplo del Teorema 1 y sus implicancias según nuestra teoría puede verse en la Figura 1.

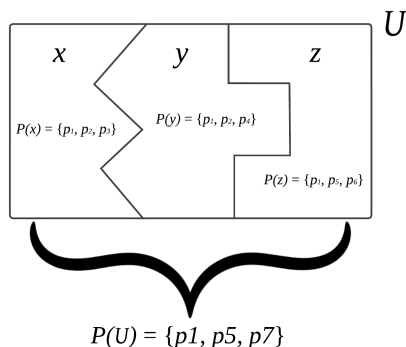


Figura 5: Ejemplificación del Teorema 1 y sus implicaciones. Donde  $U = x \dagger y \dagger z$ . Con  $P(x) = \{p_1, p_2, p_3\}$ ,  $P(y) = \{p_1, p_2, p_4\}$ ,  $P(z) = \{p_1, p_5, p_6\}$  cuyo resultante es  $P(U) = \{p_1, p_5, p_7\}$ . (Creación propia).

Como se ilustra en esta figura, cada objeto del universo —incluyendo al universo mismo— posee un conjunto único de propiedades. Aunque algunos objetos pueden compartir ciertas propiedades entre sí, esto no contradice la unicidad de sus conjuntos de propiedades, siempre que no compartan la totalidad de sus elementos. De hecho, como se establece en el Postulado 1, todos los entes comparten al menos una propiedad fundamental: la mutabilidad.

Nótese también que la suma mereológica de entidades no implica —al menos no necesariamente— que su entidad resultante adquiera todas las propiedades de sus partes constituyentes. Si ese fuera el caso, el universo tendría todos los elementos de  $P$ , cosa que es absurda a

razón de que existen propiedades contradictorias o no compatibles entre sí, como la liquidez y la solidez o tener un espín entero y un espín semientero.

### 3. Sobre el cambio: las entrañas del tiempo

A partir de lo expuesto hasta ahora, es necesario profundizar en el concepto de cambiabilidad o mutabilidad a razón de que una teoría relacional del tiempo exige el uso de este concepto para explicar adecuadamente la naturaleza de lo que denominamos tiempo.

De manera intuitiva, podemos estar de acuerdo en que la mutabilidad se refiere a la capacidad o propiedad de experimentar cambios. No obstante, aunque esta noción es ampliamente aceptada, carece de precisión conceptual. Los antiguos reconocieron la existencia de lo que denominamos cambio y dedicaron parte de su labor intelectual a esclarecer su naturaleza. Tanto Platón como Aristóteles, por ejemplo, defendieron el cambio como un fenómeno real inherente a las cosas del mundo, abordándolo en el marco de sus respectivas teorías sobre la naturaleza. Por su parte, los filósofos eleáticos, en especial Parménides y Zenón, sostuvieron que el cambio es meramente ilusorio, pues afirmaban que el mundo es inmutable y permanece en un estado estático (cf. Copleston 1969, pp. 51-72, 248-256, 321-331).

Para gran parte de la filosofía medieval, la naturaleza de lo mutable se analizaba en contraposición a la naturaleza divina, concebida como eterna e inmutable. En este contexto, lo mundano se entendía como esencialmente sujeto al cambio. Dado que el tiempo está intrínsecamente vinculado a lo mutable, se concluye que Dios, al ser inmutable, debe existir fuera del ámbito temporal (cf. por ejemplo Agustín de Hipona 2010, pp. 543-583; Boecio 141-147; Tomás de Aquino 1988, pp. 149-157).

En la modernidad, el punto de inflexión lo marcó Newton con sus *Principia*, donde ofrece dos definiciones de tiempo. La primera definición describe un tiempo de inherente atractivo científico, pues es *absoluto, verdadero y matemático, en sí mismo y por su propia naturaleza, sin referencia a nada externo* (Newton 2004, p. 64). La segunda, de uso vulgar según dice Newton, corresponde a una idea del tiempo dada a los sentidos en función del movimiento de las cosas, o sea, un tiempo relacional. Evidentemente, para Newton, tiempo y cambio estaban relacionados, pero en su concepción el tiempo es ontológicamente independiente del cambio de las cosas. Esta idea fue ampliamente criticada



por Leibniz, el primer defensor acérrimo del relacionismo respecto al tiempo y al espacio, quien mencionó en contra de Newton que: *todo lo que existe es simultáneo con otras existencias, o anterior o posterior. El tiempo es el orden de existencia de aquellas cosas que no son simultáneas* (1956, pp. 1083).

Posteriormente, la ciencia contemporánea fortaleció al menos en el plano discursivo la noción sustantivalista, lo que llevó a que el estudio del cambio, en cierto modo separado de la investigación sobre la naturaleza del tiempo, se enfocara principalmente en la causalidad y las leyes naturales, entre otros factores. En lo que a este trabajo respecta, la noción de cambio o mutabilidad se especificará con la siguiente definición:

**Definición 1:** Un objeto  $x$  es mutable si y solo si existe un instante  $t_n$  y un instante  $t_i$  posterior a  $t_n$  tal que  $P(x)$  en  $t_n$  es distinto de  $P(x)$  en  $t_i$ .

$$C(x) \Leftrightarrow \exists t_n \exists t_i > t_n (P(x)_{t_n} \neq P(x)_{t_i})$$

Donde  $P(x)_{t_n}$  corresponde al estado de  $x$  en el instante  $t_n$  y  $P(x)_{t_i}$  es el estado de  $x$  en el instante  $t_i$  posterior a  $t_n$ .

Provisionalmente, esta definición es compatible con la idea tradicional de suceso o evento como cambios de estados en una cierta cosa. Nótese que identificar el concepto de evento como “algo que acaece en un cierto período de tiempo” o expresiones afines sería inadecuado para nuestro objetivo, puesto que aún no contamos con una definición propia y precisa del concepto de tiempo.

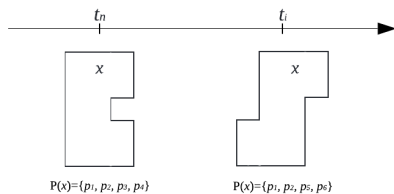


Figura 6: Ejemplo del cambio de una cierta cosa  $x$  cuyo estado inicial está en  $t_n$  y su estado final en  $t_i$ . (Creación propia).

Con esto, podemos afirmar que, por ejemplo, un cuerpo de agua ha pasado de un estado líquido a uno sólido, o que un edificio ha pasado de estar en pie a estar derrumbado. Sin embargo, este análisis puede

abordarse desde la perspectiva de las paradojas de Zenón. Según dicha óptica, entre el estado líquido inicial del agua y su estado sólido posterior, existiría un estado intermedio en el que las partículas del agua presentan una energía cinética menor que en su fase líquida pero mayor que en su estado sólido. Si seguimos el razonamiento de Zenón, podríamos postular una serie divisible *ad infinitum* y, en consecuencia, negar la existencia del cambio; no obstante, este no es el enfoque que adoptaremos.

Por ello, seguiremos caracterizando nuestra noción de cambio con la

**Definición 2:** Denominaremos devenir de un objeto  $x$  a la sucesión (o secuencia) de todos sus estados dentro de un intervalo delimitado por dos instantes. Dicho de otro modo, dado el intervalo  $[t_n - t_i]$  el devenir de  $x$  se representa como una  $n$ -tupla que contiene todos los estados de  $x$  en dicho intervalo:

$$\Delta(x)_{[t_n-t_i]} \equiv \langle P(x)_{t_n}, P(x)_{t_{n+1}}, \dots, P(x)_{t_i} \rangle$$

o de forma equivalente

$$\Delta(x)_{[t_n-t_i]} \equiv \langle P(x)_{t_k} \rangle_{k=n}^i$$

Donde  $\Delta(x)_{[t_n-t_i]}$  es el devenir de la cosa  $x$  en el intervalo de instantes  $t_n$  y  $t_i$ , y  $\langle P(x)_{t_k} \rangle_{k=n}^i$  es la sucesión de estados de  $x$  en instantes indexados por  $k$  con  $k \in \langle n, n + 1, \dots, i \rangle$ .

Añadiremos también el

**Postulado 5:** Para cualquier objeto existe un cambio mínimo, i.e., un intervalo de estados tal que no hay ningún otro estado intermedio entre ellos.

Este postulado explicita lo que en la Definición 2 está implícito: la mutabilidad ocurre de manera discreta. Esto es así porque una ontología basada en cambios continuos introduce mayores complejidades al intentar justificar el fenómeno del cambio y, en consecuencia, la noción de tiempo. La no discretización del fenómeno de cambio presenta complicaciones para, por ejemplo, responder qué ocurre con la energía en una secuencia infinita de cambios.

Podemos formalizar esta idea con la siguiente definición.

**Definición 3:** Llamaremos cambio mínimo o cambio atómico a un intervalo de estados de un objeto  $x$  tal que el devenir de  $x$  en dicho intervalo es una 2-tupla. Esto se puede expresar de la siguiente manera:

$\Delta(x)_{[t_n-t_i]}$  es un cambio mínimo  $\Leftrightarrow \Delta(x)_{[t_n-t_i]} = \langle P(x)_{t_n}, P(x)_{t_i} \rangle$

#### 4. Causalidad y modalidad desde una perspectiva relacional

Como nuestra teoría es tributaria de la teoría causal del tiempo, debemos encargarnos de exponer el rol que desempeña la causalidad en nuestra empresa y su conexión con la mutabilidad.

La noción de causalidad está intrínsecamente ligada a la temporalidad, ya que todo evento causal conlleva la idea de una sucesión temporal. Russell, tomando como referencia el diccionario de Baldwin, define la causalidad de la siguiente manera: *Dado cualquier acontecimiento  $a_1$  existe un acontecimiento  $a_2$  y un intervalo de tiempo  $t$  tal que siempre que se produzca  $a_1$  le seguirá  $a_2$  tras un intervalo  $t$*  (2001, p. 252). De hecho, a pesar de que una definición de causalidad no mencione explícitamente el término tiempo, las relaciones causales requieren necesariamente un contexto temporal, es decir, una distinción entre lo anterior y lo posterior. En función de esto, seguiremos con el siguiente postulado:

**Postulado 6: Todo cambio tiene naturaleza causal tal que, dado un cambio de  $P(x)_{t_n}$  a  $P(x)_{t_i}$ , el estado de  $x$  en  $t_n$  es causa del estado de  $x$  en  $t_i$ . Al mismo tiempo, podemos decir que el estado de  $x$  en  $t_i$  es consecuencia del estado de  $x$  en  $t_n$ .**

$$P(x)_{t_n} \neq P(x)_{t_i} \rightarrow P(x)_{t_n} \dashv P(x)_{t_i}$$

$$\therefore P(x)_{t_i} \leftarrow P(x)_{t_n}$$

Donde “ $\dashv$ ” se interpreta como “es causa de” y “ $\leftarrow$ ” como “es consecuencia de”. Es importante subrayar que la causalidad es una relación asimétrica, es decir, posee una dirección única. Por lo tanto, ni “ $\dashv$ ” ni “ $\leftarrow$ ” son operaciones conmutativas, lo que implica que no pueden invertirse sin alterar el sentido de la relación. Esta es la razón por la cual deben leerse y entenderse conforme a lo indicado, téngase presente que en nuestra teoría  $i$  siempre es posterior a  $n$  ( $i > n$ ).

En este contexto, es posible afirmar que un estado  $a$  causó el estado  $b$ ; sin embargo, en la literatura filosófica, los términos  $a$  y  $b$  suelen ser reemplazados por conceptos que se refieren a eventos, no a cosas. Esta concepción ha sido defendida diligentemente por autores como

Russell (2001), Davidson (1995), Bunge (1997) y Kim (1993), entre otros. No obstante, en aras de la parsimonia, resulta necesario prescindir de dicha interpretación de la causalidad, ya que esta noción, además de la relación causal, debe proporcionar una explicación satisfactoria de un cambio de estado (i.e., de un evento) sin recurrir a relaciones causales, lo que complicaría innecesariamente la ontología apelando a recursos como: cambios de estado atemporales o a una suerte de nivel de la realidad donde emerge la causalidad. En consecuencia, siguiendo nuestro modelo teórico, tenemos el

**Teorema 2:** Para todo proceso de cambio existen causas mínimas o causas atómicas, i.e., la relación causal que vincula los cambios mínimos existentes en dicho proceso de cambio:

$$\Delta(x)_{[t_n-t_i]} \text{ es un cambio mínimo} \rightarrow P(x)_{t_n} \Rightarrow P(x)_{t_i}$$

$$\therefore P(x)_{t_i} \Leftarrow P(x)_{t_n}$$

Donde “ $\Rightarrow$ ” se lee “es causa mínima” y “ $\Leftarrow$ ” “es consecuencia mínima”.

**Demostración:** Dado que según el Postulado 5 los cambios están discretizados en cambios mínimos y de acuerdo con el Postulado 6 todo cambio posee una naturaleza causal, se sigue que los cambios mínimos están conectados por una causa mínima. Por lo tanto, cada estado en un proceso de cambio está relacionado causalmente con el estado posterior inmediato.

**Corolario:** La causalidad se da entre los estados de una cosa, no entre eventos.

**Ejemplo:** Si  $\Delta(x)_{[t_n-t_i]} = \langle P(x)_{t_n}, P(x)_{t_k}, P(x)_{t_m}, P(x)_{t_i} \rangle$ , entonces:  $(P(x)_{t_n} \Rightarrow P(x)_{t_k})$ ,  $(P(x)_{t_k} \Rightarrow P(x)_{t_m})$ ,  $(P(x)_{t_m} \Rightarrow P(x)_{t_i})$ .

Hasta este punto, la causalidad no ha sido formalmente definida y la propuesta según la cual las relaciones causales se establecen entre estados sucesivos de un mismo objeto puede resultar poco intuitiva. Usualmente, se concibe la causalidad como una relación que involucra al menos dos objetos distintos. Un ejemplo clásico sería: el movimiento de la bola  $x$  al impactar a la bola  $y$  causa el movimiento de la bola  $y$ . No obstante, un proceso como este puede ser descrito rigurosamente dentro del marco teórico que hemos propuesto. Por ejemplo, dado un instante  $t_n$  en el que  $x$  inicia su movimiento a una distancia  $d$  de  $y$  y un ulterior instante  $t_i$  en el que  $y$  está en movimiento causado por el impacto de  $x$  (denotemos este sistema como “ $\sigma$ ”), entonces en el

intervalo  $[t_n - t_i]$  existe una cantidad finita de estados sucesivos en los que  $\sigma$  cambia hasta alcanzar  $P(\sigma)_{t_i}$  tal como muestra la Figura 5.

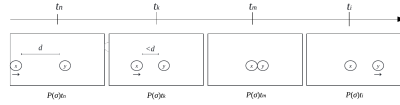


Figura 7: Reinterpretación del concepto tradicional de causalidad con base en nuestra teoría donde  $P(\sigma)_{t_n} \Rightarrow P(\sigma)_{t_k} \Rightarrow P(\sigma)_{t_m} \Rightarrow P(\sigma)_{t_i}$ .

Además, entre otras propiedades  $\Delta(\sigma)$  se puede caracterizar porque en cada uno de sus estados sucesivos  $x$  e  $y$  se encuentran a una distancia menor que en el estado anterior hasta alcanzar un estado  $P(\sigma)_{t_m}$  en el que la distancia entre  $x$  e  $y$  es 0, momento en el cual  $y$  inicia su movimiento. Asimismo, puesto que  $(x \wedge y) \sqsubset \sigma$ , podemos afirmar que en cada uno de los instantes del intervalo  $[t_n - t_i]$  existen simultáneamente (o coexisten)  $P(\sigma)$ ,  $P(x)$  y  $P(y)$ .

Sumamos también el siguiente teorema:

**Teorema 3: El devenir de cualquier ente en cualquier intervalo de instantes es único; tal que:**

$$\forall x (\Delta(x)_{[t_n - t_i]} \text{ es } \acute{u}\text{nico})$$

**Demostración:** Dado el Postulado 3, cada ente está individuado por un conjunto único de propiedades. Si los estados de un ente son conjuntos de propiedades, entonces el devenir de un ente es una sucesión única de estados. Como consecuencia, el devenir de cualquier ente no puede coincidir con el devenir de otro ente, ya que la identidad de estos depende de sus propiedades únicas.

**Corolario:**  $\Delta(U)_{[t_n - t_i]}$  es único.

Consecuentemente tenemos otro teorema:

**Teorema 4: El cambio en cualquier ente afecta y determina el cambio del universo y, en consecuencia, determina su devenir.**

$$\begin{aligned} \forall x \sqsubset U (P(x)_{t_n} \neq P(x)_{t_i} \rightarrow P(U)_{t_n} \neq P(U)_{t_i}) \\ \therefore P(x)_{t_n} \neq P(x)_{t_i} \text{ determina } \Delta(U)_{[t_n - t_i]} \end{aligned}$$

**Demostración:** Según el Corolario 2 del Postulado 4, no existen cosas aisladas y todo ente está asociado al menos con alguna otra cosa.

Esto implica que cualquier cambio en las propiedades de un ente afecta al ente con el que está asociado y, por transitividad, puede afectar a todo el conjunto de entes que forman el universo. Como el universo es la suma mereológica de todos los entes, cualquier cambio mínimo en un ente tiene alguna repercusión en el devenir global del universo.

**Corolario:** No hay cambios independientes en el universo; cada cambio, por pequeño que sea, tiene implicaciones en el sistema global.

Para profundizar en nuestra noción de causalidad, proponemos el

**Postulado 7: Todos los entes en virtud de sus propiedades tienen naturaleza modal.**

Para ser más precisos, conviene explicitar la siguiente definición:

**Definición 4: Un ente  $x$  es modal si y solo si dado el conjunto  $S = \{s_1, \dots, s_n\}$  de todos los estados posibles (i.e., de todas las combinaciones posibles de propiedades compatibles entre sí) existe un  $S' \subset S$  tal que  $S'$  constituye la totalidad del espacio de estados posibles para  $x$  con  $S' \neq \emptyset$ . O de manera equivalente:**

$x$  es modal  $\Leftrightarrow \exists S' \subset S, S' \neq \emptyset : S'$  es el espacio de estados posibles de  $x$

Por lo tanto, de acuerdo con la Definición 4, todos los entes deben necesariamente cambiar hacia un estado que les es objetivamente posible. Este conjunto de estados posibles está determinado por las propiedades que dichos entes poseen. Es importante señalar que si bien  $S'$  puede contener más de un elemento, no todos los estados son necesariamente equiprobables; es decir,  $x$  no tiene necesariamente la misma probabilidad de asumir cualquiera de los estados pertenecientes a  $S'$ .

**Ejemplo:** Si  $x$  posee el conjunto de propiedades  $\{p_1, p_2\}$ , entonces  $x$  tiene el espacio de estados posibles  $\{\{p_1, p_3\}, \{p_1, p_4\}, \{p_1, p_4, p_5\}\}$ .

Es evidente que nuestra propuesta guarda grandes similitudes con el disposicionalismo, de hecho, podríamos afirmar que nuestra ontología en algún sentido es disposicionalista en vista de que en esta teoría las propiedades y la modalidad ocupan un lugar preponderante. En palabras de Bird “las leyes refieren a las propiedades de las cosas” (2007, p. 1) y “las leyes surgen del interior de las propiedades mismas” (ibid. 2). Así, el disposicionalismo propone que las propiedades son modales per se en vez de buscar la modalidad en entidades abstractas e impropias de los entes. Esto marca una clara distancia con teorías nomológicas como la de Lewis (1986; 1994) o como la de Armstrong (1983). No obstante, como no tenemos aún una definición de ley, no

podemos decir que nuestra ontología es propiamente disposicionalista en el sentido que plantea Bird.

Ahora bien, podría argumentarse que en aras de preservar un minimalismo ontológico sería preferible adoptar una concepción humeana del mundo, es decir, amodal. Sin embargo, dicha concepción no se ajusta a la realidad observada, ya que en un mundo amodal el conjunto  $S$  representaría la totalidad del espacio de estados posibles de todos los entes, de modo que para cualquier objeto  $x$  cada elemento de  $S$  sería equiprobable. En última instancia, la amodalidad no proporciona una explicación adecuada para la regularidad de los acontecimientos en el mundo. Por consiguiente, planteamos la

**Definición 5: La causalidad es la relación modalmente establecida entre un estado anterior y uno posterior de un ente.**

Esta forma de concebir la causalidad es particularmente hábil para describir el cambio del universo en su conjunto, contrario a la concepción tradicional de causalidad entre eventos que para tales fines resulta menos parsimoniosa.

Para finalizar, ofreceremos dos definiciones más con las que podemos completar los fundamentos mínimos de una ontología científica basada en un tiempo relacional.

**Definición 6: Se denomina ley al hecho de que ciertos conjuntos de propiedades en una sucesión de estados se repitan con regularidad en otras sucesiones de estados bajo condiciones similares. Dicho de otro modo, una ley es un concepto que denota la repetición de determinadas propiedades generada por la naturaleza modal del mundo.**

Por consiguiente, decimos que hay ley cuando, por ejemplo, dado un devenir  $\Delta(x)_{[t_n-t_k]}$  en el que  $P(x)_{t_k} = \alpha = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}\}$  y un devenir posterior  $\Delta(y)_{[t_m-t_i]}$  bajo condiciones similares a las de  $\Delta(x)_{[t_n-t_k]}$  tal que  $P(y)_{t_i} = \alpha' = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{11}\}$ , las propiedades de  $P(x)_{t_k}$  y  $P(y)_{t_i}$  difieren en una única propiedad, siendo evidente la repetición de las demás propiedades.

Con esta definición, el término ley no denota ninguna entidad particular del mundo ni la naturaleza modal del mismo. En este sentido, una ley no produce regularidades, pues las regularidades son producidas por las relaciones causales intrínsecamente modales. Este concepto de ley es similar en algunos aspectos al planteado por Bunge (1977) y Romero (2018); similar en tanto las propiedades repetidas son el elemento central, pero diferente en tanto nuestro concepto de ley no refiere a la acción modal de los entes, sino a un resultado de esta.

**Definición 7:** El tiempo es la herramienta epistémica con la que cuantificamos el devenir del universo en cualquier intervalo de instantes; por tanto:

“Tiempo”  $\equiv$  la cuantificación de  $\Delta(U)_{[t_n-t_i]} \equiv$  la cuantificación de  $\langle P(U)_{t_k} \rangle_{k=n}^i$

## 5. Conclusiones

En este trabajo, hemos desarrollado una propuesta ontológica basada en el relacionismo respecto al tiempo, con objeto de aportar un marco teórico mínimo con el que se pueda construir una ontología científica minimalista. Hemos desarrollado nuestra teoría siguiendo una forma axiomática, de tal modo que evitamos la vaguedad de sus conceptos fundamentales y procuramos la solidez de sus argumentos.

Con todo, esta no deja de ser una propuesta guía, en el sentido de que expone lo mínimo necesario para un posterior desarrollo acabado de la mejor ontología científica (en nuestro caso, mejor equivale a minimalista o parsimoniosa). Con ello, queda bastante reforzada la tesis relacionista, puesto que hemos mostrado que concebir el tiempo como sustancia es algo prescindible de cara al entendimiento científico del mundo.

Hay conceptos propios de la ontología y de la metafísica del tiempo que ameritan un examen propio; por ende, eventualmente algunos de nuestros postulados podrían cambiar en virtud de aquello. Por ejemplo, el concepto de presente no aparece en nuestra teoría; así quedan fuera de este trabajo, y por ende disponibles para ampliar la discusión, conceptos como *tensed theory of time*, presentismo o eternalismo. No obstante, hemos articulado un sistema teórico tal que sirve como la base para discutir mentados conceptos.

Tanto un defensor del eternalismo como del presentismo podrá estar de acuerdo con nuestra Definición 2. El primero, sin embargo, asegurará que  $\langle P(x)_{t_n}, P(x)_{t_{n+1}}, \dots, P(x)_{t_i} \rangle$  es una secuencia de estados coexistentes. El segundo, por su parte, defenderá que es una sucesión de estados de los cuales sólo existe uno a la vez. Asimismo, un defensor de la continuidad espaciotemporal podrá estar de acuerdo en la mayor parte de lo planteado, con la salvedad de que definiría el devenir como cambios de estados continuos y modificaría nuestra formalización utilizando, por ejemplo, la noción de límite matemático. Todo esto no hace sino mostrar la preeminencia de la tesis relacional, que consigue ser más parsimoniosa que su teoría rival.



## Referencias

- Armstrong, David Malet. (1983). *What is a Law of Nature?*. Cambridge Philosophy Classics.
- Barbour, Julian B. & Bertotti, Bruno. (1982). Mach's principle and the structure of dynamical theories. *Proceedings of the Royal Society London*, pp. 295-306.
- Barbour, Julian B. (1982). Relational concepts of space and time. *British Journal for the Philosophy of Science*, 33 (3), pp. 251-274.
- Barbour, Julian B. (1986). Leibnizian time, Machian dynamics and quantum gravity. En Roger Penrose & C. J. Isham (eds.) *Quantum concepts in space and time*, pp. 236-246. New York: Oxford University Press.
- Barbour, Julian B. (1999). *The End of Time: The Next Revolution in Physics*. Weidenfeld & Nicholson.
- Bird, Alexander. (2007). *Nature's metaphysics: laws and properties*. Clarendon.
- Boecio, Severino Anicio Manilo. (2018). *La consolación de la filosofía*. Madrid: Editorial Verbum.
- Bunge, M. & Máynez, A.G. (1976). A relational theory of physical space. *Int J Theor Phys* 15 , pp. 961–972. <https://doi.org/10.1007/BF01807716>
- Bunge, Mario. (1977). *Ontología I: El Moblaje del Mundo: Volumen III. Tratado de Filosofía*. Barcelona: Editorial Gedisa.
- Bunge, Mario. (1997). *La causalidad: el principio de causalidad en la ciencia moderna*. Buenos Aires: Editorial Sudamericana.
- Copleston, Frederick. (1969). *Historia de la Filosofía Vol.1 Grecia y Roma*. Barcelona: Editorial Ariel.
- Davidson, Donald. (1995). Causal Relations. *Essays on Actions and Events*, pp. 149-162.
- de Aquino, Tomás. (1988). *Suma de Teología*. Madrid: Biblioteca de Autores Cristianos.

- de Hipona, Agustín. (2010). *Confesiones*. Madrid: Editorial Gredos.
- Hawley, Katherine. (2009). Metaphysics and Relativity. En Robin Le Poidevin, Peter Simons, Andrew McGonigal, Ross P. Cameron (eds.) *The Routledge Companion to Metaphysics*, pp. 507-516. New York: Routledge.
- Kim, Jaegwon. (1993). *Supervenience and mind*. Cambridge University Press.
- Leibniz, Gottfried. (1956). *Philosophical Papers and Letters*. Chicago: University of Chicago Press.
- Lewis, David. (1986). *Philosophical Papers vol. II*. Oxford University Press.
- Lewis, David. (1994). Humean Supervenience Debugged. *Mind* 103, pp. 473-490.
- Newton, Isaac. (2004). *The Principia*. En Andrew Janiak (ed.) *Philosophical Writings*, pp. 40-93. Cambridge University Press.
- Rovelli, Carlo. (2018). *El orden del tiempo*. Anagrama.
- Russell, Bertrand. (2001). Sobre la noción de causa. *Misticismo y Lógica*, pp. 247-284. Editorial Edhasa.

# El empréstito como herramienta de entrecruce en *Maniac*: entre la ciencia y la literatura

Mathias Riquelme Z.<sup>a</sup>

## Resumen

A veces la ciencia se muestra como un cuerpo de ideas racionales que busca una comprensión profunda del mundo. A la literatura también se le ve como el arte de explorar lo oscuro y peligroso. En estas definiciones tanto la ciencia como la literatura buscan explorar y revelar aspectos profundos de la realidad haciendo uso de diferentes métodos. En la novela de Benjamín Labatut *Maniac* se emplea un enfoque que mezcla ciencia hechos y ficción haciendo uso del recurso literario del empréstito para contar historias sobre avances científicos y figuras históricas. Este último entendido como el préstamo de elementos narrativos de una disciplina para enriquecer otra será el pináculo de toda la obra donde el autor entrelaza conceptos científicos con relatos ficcionales para explorar la condición humana y el impacto de la ciencia en el siglo XX y XXI. En el presente artículo buscamos analizar el uso del empréstito en un texto como *Maniac* y de qué manera se puede determinar la veracidad de los discursos ficcionales que incorporan hechos reales. A través de trabajos como el de John Searle se discutirá el estatus lógico del discurso de la ficción sugiriendo cómo en estas obras la literatura puede ofrecer una perspectiva única sobre la ciencia haciendo el conocimiento más accesible y reflexivo.

**Palabras claves:** Ficción, prestación, recursos narrativos, investigación, discursos ficcionales.

---

<sup>a</sup>Pedagogía en Filosofía, Universidad de Valparaíso.  
Contacto: mathias.riquelme@alumnos.uv.cl

## Abstract

Science is sometimes portrayed as a body of rational ideas that seeks a deep understanding of the world. Literature is also seen as the art of exploring the dark and dangerous. In these definitions both science and literature seek to explore and reveal deep aspects of reality using different methods. In Benjamin Labatut's novel *Maniac* an approach is used that mixes science fact and fiction making use of the literary resource of borrowing to tell stories about scientific advances and historical figures. The latter understood as the borrowing of narrative elements from one discipline to enrich another will be the high point of the entire work where the author intertwines scientific concepts with fictional relationships to explore the human condition and the impact of science in the 20th and 21st centuries. In this article we seek to analyze the use of borrowing in a text such as *Maniac* and how the veracity of fictional discourses that incorporate real facts can be determined. Through works such as those of John Searle the logical status of fictional discourse will be discussed suggesting how in these works literature can offer a unique perspective on science making knowledge more accessible and reflective.

**Key words:** Fiction, borrowing, narrative resources, research, fictional discourses.

## 1. Introducción

Para Mario Bunge, filósofo y físico argentino, la ciencia es “aquel creciente cuerpo de ideas racional, sistemático, exacto, del cual, por medio de su investigación, el hombre ha alcanzado una reconstrucción conceptual del mundo que es cada vez más profunda” [1]. Para Roberto Bolaño, escritor y literato chileno, la literatura es ante todo “saber meter la cabeza en lo oscuro, saber saltar al vacío, saber que se trata de un oficio peligroso” [5]. Así mismo, para Benjamín Labatut, periodista y escritor chileno, la literatura “no se trataría nunca de un acto de reconocimiento, sino que siempre uno de descubrimiento” [9] en su significado más primitivo.

Los primeros dos alcances entregados anteriormente nos brindan aspectos esenciales de lo que es una y la otra disciplina, encargada una de procurar una investigación del mundo, mientras que la otra procura señalar los recovecos de un acto tan básico como el escribir. Sin embargo, la última definición tiene una especie de esencialidad que parece converger a las dos anteriores: podría decirse que la ciencia, al igual que la literatura, representa en sí misma un acto de meter la cabeza en lo oscuro, de allanar una porción problemática de la realidad con un método, técnicas y, sobre todo, buscar des-cubrirlas. Encontramos

de esta forma un entrecruce entre ambas disciplinas, elementos generales que pueden resultar comunes y que permitirían el enriquecimiento mutuo.

Esta “prestación” entre la ciencia y la literatura es lo que nos interesa y, sobre todo, cuál vendría siendo el estatus lógico que tiene un relato sobre un hecho científico comprobado pero contado a partir del discurso ficcional de la literatura; abordar en última instancia hasta qué punto podemos hacer uso de la herramienta narrativa del empréstito. De esta forma, en el siguiente artículo queremos mostrar, a partir de la novela de Benjamín Labatut *Maniac*, que ella representa el uso en sí mismo de este recurso a la hora de crear una obra de ficción. Para esto, partiremos desde el examen del entrecruce narrativo generado en la obra entre dos de sus puntos centrales: lo fáctico histórico y científico, y la creación de los relatos ficcionales. Luego, y contando igualmente como antecedente con el ensayo del filósofo estadounidense John Searle *El estatus lógico del discurso de ficción* y el artículo del filósofo suizo Roman Frigg *Modelos y Ficción*, demostraremos entonces la importancia del empréstito en la relación existente entre las verdades de la ciencia, su trabajo con la ficción y el uso de esta como constitutivo de la creación de una novela como *Maniac*.

## 2. *Maniac* y el empréstito

*Maniac*, desde su lanzamiento en octubre del 2023, se ha enfrentado a múltiples problemas concernientes a su categorización como obra literaria. Para algunos, se trata de una novela, para otros, un ensayo, y para un grupo no menor, cabe dentro de una ficción histórica. Lo cierto es que el más reciente trabajo de Labatut presenta esta equivocidad netamente debido a su contenido, el cual relata los avances y límites de la ciencia en pleno siglo XX y siglo XXI de la mano de las figuras relevantes que crearon estas posibilidades. Así, en la obra desfilan personajes como el físico Paul Ehrenfest, lógicos como Georg Cantor y “genios” como el conocido John von Neumann, todo de la mano del relato literario. Labatut lo que pretende en la obra es presentar acercamientos a los distintos fenómenos científicos que han impactado el mundo contemporáneo, explicando teorías y realizando conexiones, sin por ello alejarse del propósito ficcional que tiene el texto.

Paradójicamente, si bien es manifiesta esta discusión sobre la clasificación del texto, es a simple vista que el lector puede darse cuenta de cuáles son los objetivos del autor. Sí, estamos tratando con ideas de

grueso calibre como el Teorema de Gödel y la Teoría de los organismos auto replicantes, sin embargo, no estamos buscando exponer paso a paso los cálculos y axiomas que las hacen posibles. Labatut lo que hace en última instancia es presentar estas ideas con un propósito mucho más específico, el cual es fortalecer el relato literario ficcional que se crea a partir de la vida de los científicos y cómo esta impactó en su ciencia. Caso ejemplar es la forma de tratar la vida y obra del matemático húngaro-estadounidense John Von Neumann, del cual versa la mayor parte del texto: a partir de un relato coral ficticio que Labatut crea sobre las diversas personas circundantes a von Neumann, nos hacemos una idea de su vida, trabajo e impacto en el desarrollo de la ciencia contemporánea.

Ahora bien, es aquí donde comenzamos a hacer manifiesto el entrecruce entre las distintas disciplinas. A menudo, tanto en entrevistas como en distintas ponencias, se habla del trabajo del autor como aquel que muestra “cuando la ciencia se convierte en literatura” [6], frase que busca encarnar una relación presente en *Maniac*, a saber, el que podemos utilizar elementos de una disciplina para contar historias en otra. Tal como Bolaño hace uso de la historia real de Carlos Wieder para relatar *Estrella Distante*, es posible realizar una prestación de la historia y descubrimientos de la ciencia para hablar literariamente de esto mismo con una significación diferente. La obra de Benjamín Labatut consta de la ciencia en sí misma, se abordan los conceptos, teorías y temas que pertenecen a los grandes libros y ensayos de los científicos que los hicieron, no obstante, desde la vereda de la literatura.

*Maniac* hace patente la utilización de un recurso narrativo llamado empréstito, el cual en el mundo específico de la literatura carece de definiciones exactas pero que podría entenderse bajo su acepción en economía. Según esta última, un empréstito corresponde a “un intercambio en el que una entidad de crédito emite un conjunto de valores en serie que la gente puede comprar para, a cambio, comprometerse la entidad a devolver el dinero en el plazo y con los intereses pactados” [3]. Ahora bien, en literatura el uso del concepto parte desde esta esencialidad de la idea de un intercambio de valores, mas en este caso, elementos y herramientas literarias que pueden ser utilizadas en otro discurso de una manera distinta. Podemos bajo esta misma idea hacer un préstamo de maneras de contar las cosas que tiene por ejemplo la literatura para hacer un relato que explique la ciencia o la historia. A fin de cuentas, el uso de la idea de empréstito en literatura representa

en sí mismo el acto que pretende describir, intercambiando así su uso desde la economía.

Es de esta forma que encontramos cómo la base narrativa de la obra es a partir del uso del empréstito, reconociendo el intercambio de “la manera de escribir” de las novelas históricas y el contenido de la investigación científica del siglo XX, con el propósito de algo más grande. La ciencia, conocida por su predominante rol en nuestras sociedades y el progreso de ellas, adquiere entonces un nuevo sentido, uno de explicitación distinto al común y que se logra materializar a partir de la literatura y la ficción. Labatut en ningún momento niega la necesidad de crear diálogos, frases y hasta “testimonios” que corresponderían a los distintos científicos, sin embargo, el uso de la ficción es con el propósito de sustentar la misma divulgación científica. Tratándose –en palabras de la misma obra– de “un inquietante tríptico sobre los sueños del siglo XX y las pesadillas del siglo XXI” [4], el autor hace uso del empréstito con la idea de pensar la historia y avances de la ciencia en un tono distinto al habitual, adquiriendo un carácter que no le es usualmente propio en primera instancia.

### 3. El empréstito y la problemática del estatus lógico del discurso de ficción

Encontramos que la ciencia se convierte en literatura a partir del mismo anclaje entre una disciplina y otra que representa el empréstito. Ahora bien, este entrecruce viene a generar un problema el cual versa sobre las distintas maneras en que podemos llegar a caracterizar lógicamente un discurso ficcional que trate sobre elementos de la realidad.

El filósofo y matemático alemán Gottlob Frege publica en el año 1892 un importante artículo traducido a menudo como *Sobre el Sentido y la Referencia* (*Über Sinn und Bedeutung*) en el cual explicita ciertas reglas sobre el uso y la veracidad de proposiciones sobre nuestra realidad. En aquel artículo, Frege traza la diferencia entre el sentido de una oración y la referencia que esta tiene, correspondiendo la primera a la forma en que se nos presenta lo dicho en la oración, mientras que la segunda al objeto que refiere esta misma. De esta manera, proposiciones como “Sócrates es el padre de la filosofía” y “El maestro de Platón fue el mártir de la filosofía” tendrían la misma referencia, pero no el mismo sentido, puesto que si bien ambas refieren a la misma persona, estas se presentan ante nosotros con un sentido distinto.

No obstante, de acuerdo con las reglas explicitadas por Frege, se nos presentaría una dificultad en la labor de poder caracterizar las distintas proposiciones de las obras ficcionales, puesto que ello diferiría con lo establecido sobre el rol del sentido y la referencia. Proposiciones del tipo “Meursault tiene un carácter indiferente” y “El personaje principal de *El Extranjero* asesinó a alguien” serían inexactas debido a que, si bien comparten el mismo sentido, son nombres propios sin una representación en el mundo real.

Ahora bien, y de acuerdo con lo anterior, ¿qué ocurriría con el caso de *Maniac* y aquellas obras que hacen uso del empréstito? Estas obras se nos presentan de manera compleja, puesto que si bien en ellas se están relatando discursos y diálogos ficticios, los personajes que los emiten son reales, y es más, hacen referencia a sucesos que sí ocurrieron. Cuando Richard Feynman, físico teórico y personaje del libro, habla sobre el trabajo de von Neumann y la creación de la computadora MANIAC, este está comentando hechos verídicos, sin embargo, emitiendo diálogos y líneas inventadas por Labatut. ¿Cómo clasificar esto mismo? ¿Las novelas fundamentadas en el empréstito entran en un plano distinto?

El profesor y filósofo estadounidense John Searle, en su ensayo *El estatus lógico del discurso de ficción*, nos presenta dos alcances que nos permitirían poder analizar el caso de estas obras. En primer lugar, en el apartado en que refiere sobre la diferencia entre la literatura y la ficción, Searle nos dice que el hecho de que una obra sea considerada literatura viene en mayor medida determinado por los lectores, mientras que el ser ficción por el autor. Es decir, el hecho de que una obra como *Maniac* haya suscitado tanto debate respecto de su clasificación tiene que ver en primera instancia por lo que unos y otros interpretan de ella; el lector al leer la obra decide juzgarla con un ojo distinto al análisis de las aseveraciones y lo que de verdad hay en ellas, mientras que es el mismo Labatut el que desde un inicio juzga cuánto de verdad y mentira hay en el texto. Así, en una frase de la obra como “Nils Aall Barricelli exclamaba ¡Niet! Cuando un cálculo no funcionaba” [4], puede haber más ficción que realidad, debido a que aun cuando la figura de Aall Barricelli tenía un referente en el mundo real, hay un trecho muy grande entre que este efectivamente exclamaba de esa forma a que Julian Bigelow –físico y personaje en la obra– realmente haya realizado una acotación sobre ese hecho. La ficción entonces, en el relato de Labatut, ya está predeterminada por las intenciones que este mismo propuso desde la producción de la obra.



No obstante, ello aún no resolvería qué pasa con los personajes reales y sus discursos ficcionales. Para esto, Searle nos entrega diversas aseveraciones que responderían con el hecho de asumir que por más que una proposición pueda ser estipulada en el libro, el autor no tiene compromiso alguno concerniente a la verdad de aquella aseveración, siempre y cuando esto no rompa con las reglas del universo que está planteando. *Maniac* se contextualiza en gran medida en el contexto científico y político del siglo XX, refiriendo a hechos históricos como la creación de la primera bomba atómica y de las primeras computadoras; no obstante, con sus personajes reales crea diálogos y sentimientos atribuibles a ellos sin levantarse del mismo contexto en el cual se está situando todo. Una manera de distinguir la ficción de la obra con lo “verídico” de ella es, dirá Searle, “cuando lo cual con lo que se compromete el autor se aleja de las reglas de su mundo antes presentado” [8]. En *Maniac* se hace referencia a calles reales, institutos y universidades que existen, modelos de vehículos como el Ford que son reales, por lo tanto, ello es parte del mismo contexto en el cual se está situando todo. No obstante, si en algún momento se dice que John von Neumann conducía un Suzuki que podía volar, acompañado de Marilyn Monroe, estaríamos pasando al plano ficticio de la obra.

Por consiguiente, el estatus de la ficción del texto de Labatut en contraposición con lo “real” de él terminará por determinarse de acuerdo con el alcance del compromiso del autor a representar hechos reales, bien sean hechos particulares, como lo expresado por von Neumann, o por hechos generales, como si se utilizó o no una bomba atómica sobre Hiroshima en aquel mundo estipulado. Estos discursos, si bien entran en debate con lo dicho por Frege en *Sobre el Sentido y la Referencia*, adquieren el carácter ficcional a partir de lo articulado por el autor de estas obras.

#### **4. El empréstito como entrecruce de los modelos científicos y la ficción**

Ahora bien, uno de los aspectos que no hemos tratado y que puede resultar relevante como parte del uso del empréstito tiene que ver con un elemento más particular, el cual es la comparación entre los modelos en ciencia con los elementos partes de la ficción. El filósofo y profesor suizo Roman Frigg, en su artículo *Models and Fiction (Los Modelos y la Ficción)*, presenta la tesis de que hay vastos elementos en común que nos permitirían hacer una relación entre la manera en que son con-

cebidos los modelos científicos, ya sea a partir de su uso y propósitos de creación, con los elementos centrales que presenta la ficción. Entre estas dos áreas hay una brecha que el empréstito resuelve, permitiendo comprender ciertas características de los modelos como elementos ficcionales, pero sin que por ello pierdan su naturaleza de representación científica.

En primer lugar, para poder comprender mejor de qué estamos hablando, es necesaria una definición acerca de lo que es un modelo. Si bien es notoria la ambigüedad que se presenta al intentar referirnos propiamente, el filósofo argentino Alejandro Cassini entrega una definición general sobre lo que entendemos por modelo científico: “Un modelo científico es una representación idealizada de un determinado fenómeno o dominio de fenómenos” [2]. Cassini también entrega distintas características de los modelos, de la cual destacamos –como dice en la definición– su alto grado de idealización, puesto que estos se crean netamente con el propósito de representar una determinada parcela de la realidad que resulta demasiado compleja en sí misma para estudiarla tal como es. Para ello, dirá Cassini, “simplificamos y a menudo distorsionamos los fenómenos” [2] con el fin de simular cómo actúan en su estado normal, mas a partir de condiciones predeterminadas por el científico haciendo uso de la misma pragmática del modelo. Con un ejemplo que Frigg retoma de una introducción a la física, podemos querer estudiar el movimiento de una pelota de baseball lanzada por el aire, sin embargo, hay muchos factores presentes en este acto que nos dificultan la tarea. Se crea entonces un modelo que funja como una versión simplificada del fenómeno en el que las dificultades de la forma inexacta de la pelota, la resistencia del aire y la inconstancia del peso sean olvidadas, para así tener un problema cuyos factores ambientales sean posibles de tratar. Así, partimos desde la idea que los modelos son entidades creadas en las que se incluyen determinadas características que no necesariamente se corresponden con la realidad y que, por sobre todo, simulan ficcionalmente condiciones de los fenómenos.

Sobre la base de esto, Frigg en su artículo explica cómo es que distintos autores han ya caracterizado a los modelos científicos a partir de comparaciones con la ficción, utilizando ideas del concepto como la imaginación, construcción, narración, etc. Así, tenemos visiones como la de Nancy Cartwright de donde Frigg retoma que para ella “un modelo es un trabajo de ficción” [7]. Es de esta forma en que el autor irá instaurando la tesis de que el modelo científico es en sí mismo pariente de los objetos de la ficción literaria, y que tal como no hay

razones para que el lector crea que los nombres y sucesos de una novela son descripciones del mundo real, tampoco las hay para que lo que dicen los libros de textos científicos pretendan describir sistemas reales. Igualmente, se derivan otros elementos comunes como el hecho de que, al igual que en una novela, las reglas de su universo están estipuladas desde un principio y que luego a partir de ellas podemos realizar distintas inferencias sobre qué cosas ocurren en ese mundo y qué cosas son imposibles; en el modelo científico se actúa de la misma forma. Como dice Frigg, “la descripción de un sistema-modelo sólo especifica un puñado de propiedades esenciales, pero se comprende que el sistema tiene otras propiedades que las mencionadas en la descripción” [7], por lo que lo interesante del estudio del modelo científico será precisamente el ver qué implicancias hay acerca de lo que no se dice, qué ocurre con el modelo en tal situación ejemplificada y qué situaciones podemos crear con las reglas de su universo estipuladas.

Es así como encontramos el rol que tiene el empréstito en todo esto, el cual sería el causante de posibilitar aquel intercambio de propiedades. Ya vimos con *Maniac* cómo es posible estrechar lazos al contar la ciencia como literatura, sin embargo, es más clara esta relación cuando se trata de los modelos, los cuales son en sí mismos partículas específicas del trabajo científico. En la creación del modelo requerimos atender a sus características, puesto que un modelo que no pretenda ni representar ni ser idealizado difícilmente cumplirá con sus propósitos –si es que no se vuelve por ello mismo inútil–. En la creación de una novela ficcional hacemos uso de diversos elementos que pueden ser comunes o no para nosotros, los cuales son “estatizados” con el fin de que funcionen correctamente según lo que sirvan para la historia; el mundo muggle de Harry Potter puede o no saber de la existencia de la magia, pero para los fines de la historia los no-magos tienen que tomarse aquella existencia como algo normal y no crear tantos problemas sobre su procedencia.

El empréstito entonces nos permite realizar este intercambio de propiedades entre el modelo y lo ficcional, en donde hacemos uso de este último con el propósito de poder atender a los requerimientos de creación del primero. El recurso narrativo de crear un contexto, un determinado ambiente controlado en donde se desarrollan sucesos es parte de la naturaleza de los modelos y aquella característica es esencial si pretendemos conocer a partir de él.

## 5. Conclusión

La obra *Maniac* de Benjamín Labatut ejemplifica de manera contundente cómo el entrecruce narrativo entre la literatura y la ciencia puede ofrecer una perspectiva única sobre la condición humana y el mundo que habitamos. De no ser por el uso consistente del empréstito, la narrativa no tendría el impacto que tiene en la exploración de los diversos descubrimientos científicos y las vidas de quienes los hicieron, tejiendo estos elementos con una prosa evocadora de la ciencia.

El autor no se limita a describir avances científicos, sino que los presenta como metáforas vívidas de la complejidad de la experiencia humana: hace patente la necesidad de una literatura ficcional que nos ayude a darle una forma distinta al mismo relato. Labatut examina contribuciones científicas, pero también obsesiones y sus conflictos éticos pertenecientes al mismo plano de ficción que busca crear. Esto enriquece la comprensión de los desarrollos científicos, pero también profundiza en las motivaciones humanas detrás de estos avances y ayuda a explicarlos de una manera sumamente más abierta.

El entrecruce de la literatura y la ciencia en *Maniac* revela conexiones inesperadas y reveladoras entre ambos campos, sugiriendo que la creatividad y la racionalidad pueden complementarse mutuamente de maneras que desafían las fronteras convencionales del conocimiento. La literatura riesgosa de Bolaño y el enseñoramiento que pretende la ciencia de Bunge adquieren así un enfoque distinto, el cual ilumina la importancia de la interdisciplinariedad en la investigación contemporánea, recalcando cómo la narrativa y la ficción pueden ser un poderoso vehículo para explorar y comunicar ideas científicas de manera accesible.

Así, el uso del empréstito nos ayuda a poder crear textos y novelas que poseen estatus lógicos distintos, mas ayudan a la creación de un relato ficcional apropiado. *Maniac* nos celebra la fascinación por el descubrimiento científico, argumentando convincentemente que el cruce entre literatura y ciencia no solo enriquece nuestras mentes, sino que también enriquece nuestra comprensión de lo que significa ser humano en un mundo cada vez más interconectado y tecnológicamente avanzado.

## Referencias

Bunge, Mario. (1959). *La Ciencia. Su método y su filosofía*. Buenos Aires. Editorial Sudamericana.

Cassini, Alejandro. (2016). *Modelos científicos*. Diccionario Interdisciplinar Austral. [https://dia.austral.edu.ar/Modelos\\_cient](https://dia.austral.edu.ar/Modelos_cient)

*Empréstito*. (2024, 6 de julio). Wikipedia, La enciclopedia libre. <https://es.wikipedia.org/w/index.php?title=Empr%C3%A9stito&oldid=161158028>.

Labatut, Benjamín. (2023). *Maniac*. Barcelona. Editorial Anagrama.

Osorio, J.J. (2016). *La literatura para Bolaño: concepto y práctica*. Rumbos del hispanismo en el umbral del Cincuentenario de la AIH. Vol. VI. 529-535. [https://cvc.cervantes.es/literatura/aih/pdf/17/aih\\_17\\_6\\_066.pdf](https://cvc.cervantes.es/literatura/aih/pdf/17/aih_17_6_066.pdf).

Puerto de Ideas. (21 de noviembre 2022). Cuando la ciencia se convierte en literatura - Benjamín Labatut y Carmen Figueroa — Valpo 2022. YouTube. <https://www.youtube.com/watch?v=fBkoyzBw1JE>.

Roman, F. (2010). *Modelos y Ficción*. Synthese. Vol. 172 No. 2 , pp. 251-268.

Searle, J. (1974). *El estatuto lógico del discurso de ficción*. New Literary History Vol. VI , pp. 319-332. Doi: 10.2307/468422.

Vergara, C. y Retamal, P. (11 de noviembre 2023). *Como es MANIAC la novela de Benjamín Labatut*. La Tercera. <https://www.latercera.com/culto/2023/11/11/benjamin-labatut-a-lo-que-mas-temo-todos-los-dias-es-a-mi-mismo/>.



# Física y metafísica en Descartes. Una pregunta a partir de la lectura de Clarke.

Gabriel Osman Sandoval<sup>a</sup>

## Resumen

Clarke sostiene que la obra de Descartes debiese ser leída como la producción intelectual de un científico práctico que también se interesó por cuestiones metafísicas, relegando así la metafísica a una preocupación de segundo orden. Esta propuesta carga una limitación explicativa, pues es manifiesto que posterior a 1637 la publicación de la metafísica adquiere relevancia dentro de las preocupaciones intelectuales del francés, por lo que resulta complejo aproximarse a su obra en los términos propuestos por Clarke, los cuales no permiten entender por qué un científico práctico para quien la metafísica representaba una preocupación secundaria dedicó una porción importante de su tiempo a una presentación elaborada de la misma. En el presente trabajo sostengo que la importancia que adquiere la presentación de la metafísica está vinculada con la ulterior publicación de su física. Para demostrarlo revisaré el curso histórico del trabajo de Descartes, prestando especial atención al cambio estructural que afecta la presentación de su física luego de la condena de Galileo, i.e., las diferencias estructurales existentes entre *El Mundo* y los *Principios de la Filosofía*.

**Palabras clave:** Modernidad Temprana, metafísica cartesiana, aristotelismo, condena de Galileo.

## Abstract

Clarke argues that Descartes' work should be read as the intellectual production of a practicing scientist who was also interested in

---

<sup>a</sup>Licenciado en Filosofía, Universidad de Chile, Santiago, Chile.  
Contacto: gabriel.osman.sandoval01@gmail.com

metaphysical questions, relegating metaphysics to a second order concern. This proposal has an explanatory limitation, since after 1637 the publication of metaphysics acquires relevance within the intellectual concerns of the Frenchman, so it is complex to approach his work in the terms proposed by Clarke, which does not allow us to understand why a practicing scientist for whom metaphysics represented a secondary concern devoted a significant portion of his time to an elaborate presentation of metaphysics. In the present paper I argue that the importance acquired by the presentation of metaphysics is linked to the further publication of its physics. To demonstrate this, I will review the historical course of Descartes' work, paying particular attention to the structural change affecting the presentation of his physics after Galileo's condemnation, i.e., the structural differences between *The World* and *Principles of Philosophy*.

**Keywords:** Early Modernity, cartesian metaphysics, aristotelianism, Galileo's condemnation.

## 1. Consideraciones preliminares

En su reseña del libro *Descartes' Philosophy of Science* (1982) de Desmond Clarke, Gaukroger destaca, no de forma negativa, el carácter inusual del tipo de aproximación del autor hacia la obra de René Descartes, aproximación en la cual priman obras tales como los *Principios*, *El Mundo*, la *Conversación con Burman* y una porción importante de la correspondencia del francés por sobre las *Meditaciones Metafísicas*, obra que al momento de ser publicado el trabajo de Clarke gozaba de prominencia en los estudios cartesianos contemporáneos (Gaukroger 1983, p. 445). Clarke era consciente de esto. Lo hace saber en una nota tomando distancia de lo que identifica como un desbalance en la literatura cartesiana en favor de las *Meditaciones*, señalando: "Adopto aquí la aproximación opuesta. Interpreto la obra conservada de Descartes como la producción de un científico práctico, quien desgraciadamente escribió algunos breves y relativamente poco importantes ensayos filosóficos" (Clarke 1982, p. 2). Harman, que reseñó esta obra en 1984, sostiene que este es seguramente el punto de partida correcto para una exposición de la filosofía de la ciencia de Descartes (Harman 1984, p. 114). Sin embargo, a mi parecer, el tipo de lectura propuesta por Clarke en su obra no se limita únicamente a una exposición de la filosofía de la ciencia de Descartes, sino que su enfoque tiene un alcance global, es decir, abarca toda la obra del francés. Con esto quiero decir que, si bien el texto se plantea como una exposición de la filosofía de la ciencia de Descartes, las reflexiones en él expuestas son fruto de esta lectura en clave científica de la obra del francés, en donde aquellos trabajos de



índole filosófica no juegan un papel relevante y son considerados como “desgraciadamente escritos”.

La lectura de Clarke, en donde a diferencia de los estudios de sus contemporáneos, las *Meditaciones* (y con ello la metafísica cartesiana) no eran de capital importancia, encuentra sustento en pasajes en los cuales parece haber, por parte de Descartes, una postergación de su obra metafísica acompañada de una elevación de su trabajo físico. Por otro lado, también es empleado como sustento de esta lectura el carácter científico de la correspondencia de Descartes, particularmente entre 1629 y 1637 (Clarke 1982 pp. 3-4). Considerando estos puntos, Clarke sostiene:

Si la mayor parte de la obra conservada de Descartes está relacionada con cuestiones científicas (en nuestro uso común del término ciencia) y si el autor en repetidas ocasiones invita a sus lectores a considerar su obra física como su principal contribución a la posteridad, entonces no es irracional asumir que nosotros deberíamos leer su obra exactamente como él propone: como la obra de un científico práctico que también se interesó en cuestiones metodológicas, teológicas y metafísicas (Clarke 1982, p. 5).

Al aproximarse de esta manera a la obra de Descartes, las cuestiones no científicas, entre ellas la metafísica, son relegadas a preocupaciones incidentales o de segundo orden dentro de la labor intelectual del francés. Es en este punto que esta propuesta interpretativa se encuentra con una dificultad explicativa, pues es manifiesto que en los años que siguieron a la aparición del *Discurso del Método* la publicación de la metafísica, específicamente de una versión más elaborada de aquello esbozado en la cuarta parte de la obra de 1637, adquiere relevancia dentro de las preocupaciones intelectuales cartesianas, desembocando en 1641 en la publicación de las *Meditaciones*; por lo que resulta extraño aproximarse a su obra en los términos propuestos por Clarke, los cuales no permiten comprender por qué un científico práctico, para quien la metafísica representaba una preocupación incidental, dedicó una porción importante de su tiempo a una presentación elaborada de la misma.

Así las cosas, seguir la lectura propuesta por Clarke nos deja ante la siguiente disyuntiva: o bien hay razones que desconocemos que condujeron a Descartes a tomar esta determinación, o bien la metafísica no era preocupación menor dentro del quehacer intelectual del francés. Pese

a estar de acuerdo con esto último, considero que no nos deja en una mejor posición a la hora aproximarnos a la obra de Descartes, ya que, aun respondiendo de esta manera, prevalece una cuestión sin explicar, a saber: ¿por qué la publicación de la metafísica adquirió relevancia dentro de las preocupaciones intelectuales de Descartes? Esto tomando en consideración que, si bien el trabajo metafísico de Descartes se remonta a 1629, según se desprende de la carta del 15 de abril de 1630, no consideraba oportuna su publicación. Por otro lado, al embarcarse en el proyecto de publicación de su metafísica, el francés se distanció del curso que había tomado su programa intelectual en los años anteriores, el cual estaba orientado a la publicación de su obra física, i.e., *El Mundo*. Por lo tanto, independiente si se está o no de acuerdo con la lectura ofrecida por Clarke, cabe hacerse la pregunta acerca de los motivos por los cuales Descartes retorna a su trabajo metafísico de la forma en que lo hizo.

En las páginas que siguen ofreceré una respuesta a la pregunta arriba enunciada. Para ello, revisaré el curso histórico del trabajo de Descartes y el contexto en el que este es desarrollado, pues considero que un punto clave para responder a esta cuestión se encuentra en la condena de Galileo y la forma en la que esta influyó en la obra de Descartes.

## 2.

En 1629, Descartes se estableció en las Provincias Unidas. Su correspondencia sugiere que dedicó los nueve primeros meses desde su llegada a Frisia a componer un pequeño tratado de metafísica (Descartes 1964-74 I, pp. 144, 182, 350). Al tratarse de un trabajo perdido, las noticias que tenemos sobre esta obra se reducen a lo que nos informan acerca de ella un par de pasajes esparcidos en la correspondencia del autor, principalmente en los intercambios epistolares sostenidos con Marin Mersenne. Pese a sus obvias limitaciones, estos extractos permiten tener un conocimiento parcial del contenido de la obra en cuestión, el cual giraba en torno a la demostración de la existencia de Dios y la inmortalidad del alma, tópicos desarrollados con posterioridad en el *Discurso del Método*, las *Meditaciones Metafísicas* y en la parte primera de los *Principios de la Filosofía*.

Pese al tiempo que Descartes empleó en el desarrollo de esta obra y la importancia de los temas en ella abordados, sabemos que para 1630 no consideraba oportuna su presentación hasta no conocer la recepción de su obra física. En efecto, las primeras intenciones de publicación

del francés están vinculadas a sus estudios físicos. En octubre de 1629, Descartes se había propuesto poner por escrito en un pequeño tratado sus explicaciones de los fenómenos sublunares (Descartes 1964-74 I, 22). Tal proyecto rápidamente vio ampliado su alcance, y al mes siguiente el francés ya había decidido explicar todos los fenómenos de la naturaleza, i.e., toda la física (Descartes 1964-74 I, p. 70), dando inicio al desarrollo de *El Mundo o Tratado de la luz*.

Descartes trabajó en esta obra desde su concepción en 1629 hasta 1633, año en el que las intenciones de publicación fueron abruptamente abandonadas. Según se desprende de la carta dirigida por el autor a Mersenne a finales de noviembre de 1633, la condena de Galileo sería la principal razón tras esta decisión. Si bien Descartes no conocía para ese entonces el motivo de la condena, intuye que se debe a la defensa de la tesis acerca del movimiento de la Tierra por parte del astrónomo italiano, por lo que decide dar marcha atrás a la publicación de *El Mundo*. Adujo entonces que no deseaba publicar un discurso que tuviese una palabra que la Iglesia pudiese desaprobado, mostrando con ello su obediencia a la autoridad de dicha institución (Descartes 1964-74 I, p. 271, 281).

Aunque Descartes canceló la publicación de su *Mundo*, el tratado siguió formando parte de algunos intercambios epistolares. En 1637, por ejemplo, escribe a Mersenne:

Parece que teme que la publicación del *Discurso inaugural* pueda comprometerme a no publicar nunca mi Física. No debe temer eso porque en ninguna parte prometo no publicarla en vida. Simplemente digo que en el pasado planeaba publicarla, pero que más recientemente por las razones que expongo he decidido no hacerlo en vida [...] En efecto, se puede inferir de esto que si las razones que me impiden publicarla se modificaran, podría tomar una nueva resolución [...] porque cuando se elimina una causa, se elimina su efecto (Descartes 1964-74 I, p. 367).

El párrafo en cuestión sugiere que Descartes no estaba del todo cerrado a una ulterior publicación de su obra física, cuestión que también es manifestada en una carta en respuesta a un corresponsal ignoto a finales de 1637, en donde además el autor hace una interesante revelación respecto al *Discurso del Método*. Escribe:

En cuanto al tratado de física que usted ha tenido la amabilidad de instarme a publicar, no habría sido tan imprudente

de hablar de él en la forma en que lo hice si no hubiera estado deseoso de publicarlo [...] Pero quiero que sepa que la presente publicación tiene por objeto preparar el camino y tantear las aguas (Descartes 1964-74 I, p. 370).

Siguiendo lo planteado en el párrafo citado, el *Discurso* puede ser entendido como una preparación para una posterior publicación de la física cartesiana. A pesar de esto último, en los años que siguieron a la publicación de esta obra, los esfuerzos intelectuales de Descartes estuvieron, contrario a lo que uno esperaría, concentrados en la presentación de su metafísica. Por otro lado, en la correspondencia relativa a estos años no hallamos indicio alguno de que el francés planeara hacer pública su física en el corto plazo. Es más, no es sino hasta 1642, un año después de la publicación de las *Meditaciones*, que vuelve a manifestar la intención de publicar su física (Descartes 1964-74 III, pp. 523).

Volviendo sobre la última cita, cabe preguntarse de qué manera el *Discurso* prepara el camino para una ulterior publicación de *El Mundo*. Una posible respuesta se halla en la carta a Mersenne de finales de mayo de 1637, en la cual Descartes señala que si se refirió a su física de la forma en que lo hizo en el *Discurso* fue para instar a quienes quisieran leerla a poner fin a las causas que impedían su publicación (Descartes 1964-74 I p. 368). No obstante, se puede suponer que aquella preparación recae en la presentación misma que hace de su física en el *Discurso*, la que como veremos más adelante parece estar estrechamente ligada a los esfuerzos del autor por publicar su metafísica.

### 3.

A diferencia de *El Mundo*, que si bien contiene algunas consideraciones de carácter metafísico acerca de la creación de las verdades eternas, es un texto concebido como una obra principalmente física, en el esbozo del contenido de este tratado ofrecido en la quinta parte del *Discurso*, el autor comienza haciendo hincapié en la conexión que existe entre las verdades físicas y las verdades metafísicas presentes en la cuarta parte, como si de una cadena de verdades se tratase. Escribe Descartes:

Mucho me agradaría proseguir; y mostrar aquí toda la cadena de las otras verdades que deduje de esas primeras. Pero para ello sería necesario que hablase ahora de varias cuestiones sobre las que disputan los doctos, con quienes no quisiera indisponerme (Descartes 1964-74 VI p. 40).

Esta estructura de presentación, que difiere de aquella que se encuentra presente en *El Mundo* respecto a la conexión sistemática existente entre la metafísica y la física cartesiana, es desarrollada posteriormente en los *Principios de la Filosofía*, en cuyo prefacio a la edición francesa Descartes señala:

En consecuencia, tomé el ser o la existencia de este pensamiento como mi primer principio, y desde él deduje muy claramente los siguientes principios. Hay un Dios que es el autor de todo lo que hay en el mundo; además, puesto que él es la fuente de toda verdad, ciertamente no creó en nosotros un entendimiento del tipo que sería capaz de equivocarse en sus juicios acerca de las cosas de las que posee una percepción muy clara y muy distinta. Estos son todos los principios de que me sirvo con respecto a las cosas inmateriales o metafísicas, y de ellos deduzco muy claramente los principios de las cosas corpóreas o físicas, a saber, que hay cuerpos que se extienden a lo largo, a lo ancho y a lo profundo, y que tienen diversas formas y se mueven de diversas maneras (Descartes 1964-74 VIII B, p. 10).

El proyecto de publicación de un texto que contuviese una exposición sistemática de su filosofía comenzó a ser concebido por Descartes alrededor de 1640. En noviembre de dicho año escribe a Mersenne: “debo decirle que he resuelto escribirlos antes de dejar ese país y publicarlos tal vez dentro de un año. Mi plan es escribir una serie de tesis las cuales constituirán un libro de texto completo de mi filosofía” (Descartes 1964-74 III, p. 233). El trabajo al que aquí se hace alusión es, por así decirlo, un germen de lo que posteriormente serán los *Principios* (1644), obra cuya estructura es bosquejada en la carta al Padre Dinet anexada en la segunda edición de las *Meditaciones* (1642) de la manera que sigue: “trataré cada tema sucesivamente en artículos breves y presentaré los temas en un orden tal que la prueba de lo que viene después dependa únicamente de lo que ha venido antes, de modo que esté conectado en única estructura” (Descartes 1964-74 VII, pp. 577). No obstante, las intenciones por parte del francés de ofrecer una presentación sistemática de su física y su metafísica comienzan a manifestarse con anterioridad; ya en mayo de 1638 declara que espera demostrar los principios de la física por medio de la metafísica (Descartes 1964-74 II, p. 141).

Lo anterior supone una clara ruptura con la estructuración de la producción intelectual de René Descartes entre los años 1629 y 1633, periodo en el cual su trabajo metafísico y su trabajo físico estaban dispuestos para ser presentados de manera separada en dos obras distintas e independientes la una de la otra, pues como mencioné anteriormente, *El Mundo* fue concebido por su autor como un texto físico en el cual la presencia de la metafísica estaba reducida a algunas consideraciones relacionadas con la creación de las verdades eternas (Descartes 1964-74 I, p. 145). Tal disposición de las materias parece sugerir una cierta desconexión entre ambas. No obstante, en la carta del 15 de abril de 1630, escribe a Mersenne:

Y estimo que todos aquellos a quienes Dios ha concedido el uso de esta razón [la razón humana] están obligados a emplearla principalmente en intentar conocerle y conocerse a sí mismos. Es por aquí por donde he tratado de comenzar mis estudios, y os diré que no habría podido encontrar los fundamentos de la física si no los hubiese buscado por esta vía. Mas es ésta la materia que más he estudiado de todas y en la cual, gracias a Dios, si no me he satisfecho en modo alguno, pienso haber encontrado por lo menos cómo se pueden demostrar las verdades metafísicas de una manera que es más evidente que las demostraciones de la geometría. Digo esto según mi juicio, pues no sé si podría persuadir de ello a los demás. Durante los nueve primeros meses que he pasado en este país, no he trabajado en otra cosa y creo que ya me habéis oído hablar anteriormente de que me había resuelto a poner algo de ello por escrito (Descartes 1964-74 I, p. 144, agregado mío, traducción de Lomba).

Este pasaje es una clara manifestación de que, a pesar de la decisión de presentar su *Tratado Metafísico* y su *Tratado Físico* por separado y la aparente desconexión que de ahí se desprende, la conexión sistemática presentada a modo de una cadena de verdades posteriormente en el *Discurso* y en los *Principios* data de un periodo temprano.

Lo que no queda claro es la razón tras la decisión por parte de Descartes de ofrecer una presentación sistemática de su pensamiento, o lo que es lo mismo, la demostración de los principios de su física mediante la metafísica, especialmente si se toma en consideración que, como fue mencionado más arriba, Descartes no juzgaba oportuno hacer pública su metafísica sin conocer la recepción de su física. Estimo que

para acercarse a una posible explicación de este fenómeno es necesario tener a la vista la manera en la que Descartes interpreta su contexto intelectual.

#### 4.

En diciembre de 1629, Descartes escribe a Mersenne que tan pronto completase un pequeño tratado en el que estaba trabajando, probablemente *el Mundo*, le haría llegar una copia, no para su publicación inmediata, sino para que la obra fuese objeto de su escrutinio y el de otras personas inteligentes (Descartes 1964-74 I , p. 85), agregando: “Deseo esto principalmente a causa de la teología, que ha estado tan subyugada a Aristóteles que es casi imposible exponer otra filosofía sin que esta parezca ir en contra de la fe” (Descartes 1964-74 I , pp. 85-86). Teniendo en consideración que Mersenne era un sacerdote perteneciente a la orden de los Mínimos, es posible que Descartes buscase que su trabajo fuese revisado por él para asegurarse de que no contuviese opinión alguna contraria a la fe, pues inmediatamente después del pasaje citado agrega: “A propósito de este asunto, por favor dígame si hay algo definido en la religión concerniente a la extensión de las cosas creadas, es decir, si esta es finita o infinita, y si en todas esas regiones llamadas espacios imaginarios hay genuinamente cuerpos creados” (Descartes 1964-74 I , p. 86).

Con razón o no, esta manera de interpretar su contexto intelectual, según la cual existe una vinculación tal entre la filosofía de Aristóteles y la teología que cualquier planteamiento que se distancie de la primera parece atentar contra la segunda, desempeña un papel importante en la confección de *El Mundo*. Desde los primeros capítulos, Descartes presenta explicaciones físicas sin recurrir a las formas substanciales aristotélicas. Véase, por ejemplo, los tratamientos de los que son objeto el calor y la luz (cap. II), así como la dureza y la liquidez (o fluidez) (cap. III), los cuales son explicados de manera mecanicista, es decir, tomando como elementos explicativos: el movimiento, el tamaño y la forma de corpúsculos que interactúan entre sí. Pero en el capítulo sexto, el autor avanza hacia una explicación mecanicista e independiente de las formas substanciales no sólo de un número limitado de fenómenos físicos sino del universo. Para esto propone a modo de fábula un relato sobre la creación de un mundo nuevo. Escribe:

Permitid pues que vuestro pensamiento durante un breve tiempo salga fuera de este mundo para acceder a otro to-

talmente nuevo, que yo haré nacer en su presencia en los espacios imaginarios. [...] adentrémonos únicamente como para perder de vista todas las criaturas que Dios hizo hace cinco o seis mil años. Y tras habernos detenido en algún lugar determinado, supongamos que Dios crea de nuevo alrededor nuestro tanta materia que hacia cualquier lado al que nuestra imaginación pueda extenderse no perciba ya ningún lugar que esté vacío (Descartes 1964-74 IX p. 32. Traducción de Rioja).

El empleo de la fábula como forma de expresión de sus planteamientos adquiere sentido cuando luego de caracterizar la materia creada por Dios en el hipotético nuevo mundo como mera extensión carente de cualidades y formas (Descartes 1964-74 IX p. 33), señala cerca del final del capítulo que su intención “no es explicar como ellos las cosas que en efecto existen en el verdadero mundo” (Descartes 1964-74 IX p. 36). Por medio de este contraste entre un mundo verdadero y un mundo ficticio, el autor evita una confrontación directa con el aristotelismo, pues, pese a que en el relato hay un claro abandono de las formas substanciales aristotélicas como elemento explicativo, siendo incluso desechadas, estas consideraciones atañen al mundo ficticio.

El relato fabulesco no se limita exclusivamente al capítulo sexto, sino que se extiende hasta el capítulo decimoquinto de manera que las consideraciones realizadas por el autor en el capítulo décimo relacionadas con el movimiento de los planetas, incluida la Tierra (Descartes 1964-74 IX, p. 64), son también parte de esta forma de presentar sus planteamientos. No obstante, pese a las consideraciones que he realizado en el párrafo anterior respecto al empleo de la fábula por parte de Descartes, el autor de igual manera decide, al enterarse de la condena de Galileo, dar marcha atrás a la publicación de esta obra. ¿Por qué? Una posible respuesta puede estribar en la sorpresa causada por el hecho de que Galileo haya sido quien fuese condenado, alguien que según Descartes entendía contaba con la buena gracia del Papa, lo que revelaría la gravedad del asunto. Por otro lado, Descartes tuvo acceso a algunas cartas acerca de la condena de Galileo impresas en Liège el 20 de septiembre de 1633, de las cuales se desprendería que incluso se prohibía plantear hipotéticamente el movimiento de la Tierra (Descartes 1964-74 I, p. 288), debido a lo cual la diferenciación entre mundo verdadero y mundo ficticio introducida por medio de la fábula no surtía efecto.



De todas maneras, cabe recalcar que en las cartas que Descartes envía a Mersenne con motivo de la condena de Galileo no se percibe algún tipo de temor por parte del primero a ser condenado. Es más, tenía conocimiento de que no toda decisión tomada por los Inquisidores Romanos era automáticamente un artículo de fe, para lo cual era necesario que fuese aprobado por un Concilio General (Descartes 1964-1974 I, p. 285). No obstante, de igual forma evitó comentar, incluso en la privacidad de su correspondencia, sus opiniones sobre el movimiento de la Tierra teniendo en consideración la posición de los censores (Descartes 1964-74 I, p. 288).

Para efectos de este escrito, lo que resulta de radical importancia es tener en cuenta que la condena de Galileo es comprendida por Descartes no como un hecho anómalo ni mucho menos aislado, sino que es interpretada, nuevamente con razón o no, como una manifestación de su contexto intelectual, llegando a escribir al respecto el 31 de marzo de 1641 que quienes estuvieron detrás de la condena del astrónomo italiano fueron aquellos que confundían a Aristóteles con la Biblia (Descartes 1964-74 III, p. 350). Es posible plantear hipotéticamente que Descartes tiene a la vista a los miembros de la Compañía de Jesús, esto por dos razones: la primera es que en 1634 Descartes subraya la importancia de los jesuitas en la condena, señalando que estos habían ayudado a que tal cosa ocurriese (Descartes 1964-74 I, p. 282); la segunda guarda relación con el vínculo de aquellos con la filosofía de Aristóteles. Cuando en 1640 Descartes decide volver a leer la filosofía jesuita, específicamente libros de texto, como una forma de prepararse para responder a sus objeciones, señala que solo recuerda algo de los *Combricences*, Toletus y Rubius (Descartes 1964-74 III, p. 185), todos ellos comentarios sobre la obra del estagirita (Lohr 1975, pp. 717-719; Lohr 1980, pp. 702-703; Lohr 1982, pp. 199-201). Razón por la cual no resulta extraño que Descartes, un exalumno de La Flèche, en donde probablemente se familiarizó con estos textos, identifique una relación filosófica entre los miembros de la Compañía y el aristotelismo.

Volviendo sobre lo planteado en la carta de marzo de 1641 (Descartes 1964-74 III p. 350), ella tiene un claro eco de la carta escrita en 1629 en donde se tematiza la relación entre la teología y la filosofía de Aristóteles (Descartes 1964-74 I, p. 86). No obstante, estos no son los únicos documentos unidos por esta temática, ya que en 1637 tiene lugar entre Descartes y el jesuita Étienne Noël un intercambio epistolar con motivo de la publicación del *Discurso*, obra de la cual el autor envió un ejemplar a Noël, quien fuese un antiguo maestro suyo en La Flèche.

Lo relevante de este intercambio se encuentra en la carta de octubre de 1637 en la cual Descartes escribe:

Y porque sé que la principal razón que hace que los vuestros rechacen muy cuidadosamente todo tipo de novedades en materia filosófica es el miedo que tienen a que no causen ellas también algún cambio en la Teología, quiero advertiros aquí particularmente que no hay absolutamente nada que temer en este respecto de las mías, y que tengo motivos para dar las gracias a Dios de que las opiniones que me han parecido las más verdaderas en la Física por la consideración de las causas naturales siempre han sido las que mejor concuerdan con los misterios de la Religión; como espero hacer ver claramente (Descartes 1964-74 I, pp. 455-456).

La relevancia de este pasaje radica principalmente en que revela algo que los otros documentos anteriormente citados no. Mientras que con la ayuda de AT I 86 y AT III 350 es posible aproximarse a una comprensión del modo en que Descartes interpreta su contexto intelectual, en la epístola a Noël se nos presenta la manera en que el francés reacciona ante este, especialmente después de la condena de Galileo. Pues si para cuando trabajaba en su *Mundo* Descartes estimaba como suficiente el recurso, o si se quiere la estrategia argumental de la fábula con sus respectivas implicancias para presentar sus planteamientos físicos, hacia 1637 estimaba oportuno e incluso necesario mostrar que estos no entraban en conflicto con la teología.

## 5.

Pese a que al referirse a los *Principios* en la carta dirigida al Padre Dinet, Descartes menciona que no presentará sus opiniones ni en el mismo orden ni estilo que lo hizo en el *Discurso* (Descartes 1964-74 VII, pp. 577), es innegable que ambas obras, como señalé anteriormente, comparten un diseño estructural en lo relativo a la presentación sistemática de la metafísica y física de Descartes, cuestión que no tiene lugar en *El Mundo*. Esta manera de estructurar sus planteamientos, en donde la física se sigue de la metafísica, está estrechamente ligada a la epístola enviada a Noël en 1637 (Descartes 1964-74 I, pp. 455-456), pues al explicitar la conexión sistemática entre ambas partes de su filosofía del modo en que se hace en estas obras, la física no entra en

conflicto con la teología, ya que en última instancia tiene como fundamento la existencia de Dios. No obstante, antes de proceder de esta manera, el francés debía demostrar que el sostén de su física, i.e., su metafísica, no era incompatible con la ortodoxia teológica, tal como hace notar Gaukroger (1995 p. 356). Es posible que esta sea la razón por la que Descartes es tan cauteloso respecto a sus planes de ofrecer una exposición sistemática de su filosofía, pues tan pronto contó sobre ellos a Mersenne, le pidió que no fuesen revelados hasta no conocer la recepción de sus *Meditaciones* (Descartes 1964-74 III, p. 233, 260).

Cuando en noviembre de 1639 Descartes comenta que consideraba presentar nuevamente su metafísica, proyecto que desembocará dos años más tarde en la publicación de las *Meditaciones*, le importaba que fuese revisada por veinte o treinta de los teólogos más eruditos a fin de recibir sus críticas y saber qué cambiar, corregir o añadir antes de la publicación (Descartes 1964-74 II, p. 622), pero la correspondencia posterior permite plantear que tal vez el interés del autor por que su obra fuese revisada por veinte o treinta de los teólogos más ilustres consistiera solamente en un paso estratégico en pos de blindar su obra ante acusaciones de heterodoxia. Pienso, por ejemplo, en la carta del 11 de noviembre de 1640, en donde Descartes escribe:

Podría contar también con la aprobación de la Sorbona, que deseo y que creo puede ser muy útil para mis propósitos, pues debo decirle que el pequeño libro de metafísica que le envié [las *Meditaciones*] contiene todos los principios de mi física (Descartes 1964-74 III, p. 233. Agregado mío).

La confesión vuelve a repetirse en la carta del 28 de enero de 1641:

Puedo decirle entre nosotros que estas seis *Meditaciones* contienen todos los fundamentos de mi física. Pero por favor no se lo diga a nadie, pues eso podría dificultar su aprobación por parte de los partidarios de Aristóteles. Espero que los lectores se acostumbrarán gradualmente a mis principios y reconocerán su verdad antes de notar que destruyen los principios de Aristóteles” (Descartes 1964-74 III, p. 298).

Por lo que, a pesar de lo expuesto por Descartes en la carta al Padre Gibieuf (Descartes 1964-74 III, pp. 237-238) y en la epístola dedicatoria a los sabios Doctores de la Sorbona anexada a las *Meditaciones* (Descartes 1964-74 VII, pp. 1-6), es posible que el interés por parte del

francés en la revisión (y protección) de su obra por parte de teólogos tuviese las intenciones anteriormente señaladas; a saber, protegerla de acusaciones de heterodoxia, legitimando así la base fundamental de su física, preparando de esta manera el camino para una ulterior publicación de esta última.

Vistas de esta manera, las *Meditaciones* parecen ser un mero texto de transición, un juicio que dada su magnitud no me atrevo a formular, pero sí sostengo que esta obra, tal como ha señalado Garber (2000, p. 223), no es un proyecto autónomo, sino que es un prelude de un proyecto científico más amplio.

## 6. Consideraciones finales

En 1930, Marc Bloch escribió: “La ciencia sólo se construye mediante aproximaciones sucesivas [...] y si todo investigador debe tomar como punto de partida la obra de sus antepasados, su primer deber hacia ella es precisamente no considerarla como un dogma” (2019, p. 247). Esta cita contiene, por así decirlo, el espíritu de este trabajo, el cual, a pesar de hacer notar lo que para quien escribe son límites explicativos que afectan a la propuesta de lectura de Clarke, no se detiene a realizar una crítica de esta posición, sino a reflexionar a partir de estas limitaciones; fue así como surgió la pregunta que dio lugar a este trabajo, a saber: ¿por qué, específicamente en los años posteriores a la aparición del *Discurso*, podemos notar que la publicación de la metafísica adquirió relevancia dentro de las preocupaciones intelectuales de Descartes? Ante lo cual podemos responder, considerando lo expuesto en las secciones previas, que dicha preocupación dista de ser autónoma y que más bien se enmarca en un proyecto de mayor envergadura cuyo propósito es en última instancia preparar el terreno para la publicación de la física cartesiana.

La condena de Galileo, entendida por Descartes como una manifestación de un contexto intelectual en el cual, dado el estrecho vínculo entre la teología y la filosofía de Aristóteles, cualquier planteamiento filosófico que se distanciara del estagirita parecía ir en contra de la fe, reveló al francés que la manera adecuada de hacer pública su física era mostrando que aquella no afectaba en forma alguna a la teología, razón por la cual, desde el *Discurso* hasta los *Principios*, la física aparece precedida por la metafísica en lo que es a todas luces un ensamblaje sistemático de ambas materias, presentadas como una cadena de verdades, siguiéndose la física de la metafísica, haciendo explícita de esta

manera la fundamentación metafísica de la física cartesiana. Esta es quizás la consecuencia más relevante de la condena de Galileo en el proyecto cartesiano; sostengo esto tomando en consideración que en los años previos a este suceso tanto la metafísica como la física estaban dispuestas para ser presentadas por su autor de manera independiente.

Por supuesto, es importante aclarar si la interpretación que Descartes hace de su contexto intelectual se condice con la realidad, pues aquello, sumado a lo escrito hasta aquí, nos entregaría, aunque de manera parcial ya que se tiene a la vista un caso particular, una imagen del desarrollo de la práctica filosófica en la así llamada Modernidad Temprana, especialmente en relación con la forma de publicación de ideas fuera del canon establecido en un contexto poco receptivo a ellas debido a una vinculación entre el canon filosófico y la teología. Espero hacerme cargo de esto en otra ocasión.

## Referencias

- Bloch, M. (2019). Fustel de Coulanges historiador de los orígenes de Francia. En *Historia e historiadores*. Traducción de Francisco Gonzáles García. Madrid: Akal.
- Clarke, D. (1982). *Descartes' Philosophy of Science*. Manchester: Pennsylvania State University Press.
- Descartes, R. (1964-1974). *Œuvres de Descartes*. Charles Adam & Paul Tannery (Eds.). París: J. Vrin.
- Descartes, R. (2011). Tres cartas a Marin Mersenne (primavera de 1630). Edición bilingüe, introducción, traducción y notas de Pedro Lomba. Madrid: Encuentro.
- Descartes, R. (2019). *El Mundo o el Tratado de la luz*. Traducción, introducción y notas de Ana Rioja Nieto. Madrid: Alianza.
- Garber, D. (2000). Semel in vita: The Scientific Background to Descartes' Meditations. En *Descartes Embodied: Reading Cartesian Philosophy through Cartesian Science* (pp. 221-256). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511605994.012>.
- Gaukroger, S. (1983). Descartes' Philosophy of Science. Desmond M. Clarke. *Isis*, 74(3), pp. 445-446. <https://doi.org/10.1086/353346>.

Gaukroger, S. (1995). *Descartes: An Intellectual Biography*. Oxford: Clarendon Press.

Harman, P. M. (1984). Seventeenth Century - Desmond M. Clarke, *Descartes' philosophy of science*. Manchester: Manchester University Press 1982. Pp. xiv + 249. ISBN 0-7190-0868-9. £19.00. *The British Journal for the History of Science*, 17(1), pp. 114–114. <https://doi.org/10.1017/S0007087400020707>.

Lohr, C. H. (1975). Renaissance Latin Aristotle Commentaries: Authors C. *Renaissance Quarterly*, 28(4), pp. 689–741. <https://doi.org/10.2307/2860175>.

Lohr, C. H. (1976). Renaissance Latin Aristotle Commentaries: Authors D-F. *Renaissance Quarterly*, 29(4), pp. 714–745. <http://www.jstor.org/stable/2860037>.

Lohr, C. H. (1980). Renaissance Latin Aristotle Commentaries: Authors Pi-Sm. *Renaissance Quarterly*, 33(4) , pp. 623–734. <https://doi.org/10.2307/2860691>.

# Revisitando el problema de Wigner: la irrazonable eficacia de las matemáticas en las ciencias naturales

Leonardo Segura<sup>a</sup>

## Resumen

Este trabajo examina la irrazonable eficacia de las matemáticas en las ciencias naturales según el influyente ensayo de Eugene Wigner, centrando la discusión en las clarificaciones contemporáneas sobre esta problemática filosófica. El análisis se concentra en cómo las matemáticas, a pesar de haber sido desarrolladas sin una finalidad física específica, resultan ser una herramienta efectiva en la formulación de teorías físicas. Además, se exponen las aclaraciones propuestas por filósofos contemporáneos de la argumentación de Wigner para evitar malentendidos en la interpretación del enigma, abordando las principales críticas y reinterpretaciones que han surgido desde la publicación original de su ensayo. En este sentido, a través del ejemplo concreto del oscilador armónico simple, se ilustra cómo los conceptos matemáticos abstractos encuentran un sitio de aplicación en el ámbito físico. El objetivo de este trabajo es ofrecer una explicación detallada y accesible de la cuestión, centrándose en destacar la importancia continua de este problema en la filosofía contemporánea y su impacto en la comprensión moderna de las ciencias naturales.

**Palabras Clave:** enigma de Wigner, aplicabilidad de las matemáticas, leyes de la naturaleza, efectividad de las matemáticas, filosofía de la ciencia.

---

<sup>a</sup>Universidad de Valparaíso, Valparaíso, Chile.  
Contacto: leonardo.seguraf@alumnos.uv.cl

## Abstract

This paper examines the unreasonable effectiveness of mathematics in the natural sciences according to Eugene Wigner's influential essay, focusing the discussion on contemporary clarifications of this philosophical problematic. The analysis concentrates on how mathematics, despite having been developed without a specific physical purpose, turns out to be an effective tool in the formulation of physical theories. In addition, the clarifications proposed by contemporary philosophers of Wigner's argumentation to avoid misunderstandings in the interpretation of the enigma are exposed, addressing the main criticisms and reinterpretations that have arisen since the original publication of his essay. In this sense, through the concrete example of the simple harmonic oscillator, it illustrates how abstract mathematical concepts find a place of application in the physical realm. The aim of the essay is to offer a detailed and accessible explanation of the issue, focusing on highlighting the continuing importance of this problem in contemporary philosophy and its impact on the modern understanding of the natural sciences.

**Keywords:** Wigner's enigma, applicability of mathematics, laws of nature, efficacy of mathematics, philosophy of science.

## 1. Introducción

Una de las preguntas cruciales al momento de desarrollar un enfoque adecuado para la modelización es entender cómo se aplican las matemáticas a los modelos. En este sentido, Frigg y Nguyen (2017, p. 54) advierten que al abordar cómo un modelo matematizado logra representar su objetivo conlleva a investigar cómo se aplican las matemáticas a un sistema físico. En efecto, no habría necesidad de separar el estudio de la cuestión de la representación del modelo del problema de la aplicabilidad de las matemáticas.

En este sentido, estudiar cómo las matemáticas se vinculan con los sistemas físicos es algo que el físico húngaro Eugene Wigner intentó dilucidar en su artículo "La irrazonable eficacia de las matemáticas en las ciencias naturales" (1960), donde plantea una cuestión intrigante en la filosofía de la ciencia: ¿por qué las matemáticas, una creación abstracta del intelecto humano, son tan efectivas para describir y predecir fenómenos naturales? Este misterio ha llamado la atención en primer lugar de científicos y, en los últimos años, de los filósofos. Wigner, un físico teórico conocido por sus contribuciones a la mecánica cuántica y la teoría de grupos, examina en su artículo varios puntos claves que abren un debate sobre la naturaleza de las matemáticas y su relación con los fenómenos físicos.



Este artículo se centra en la relación entre matemáticas y física según Wigner y examina cómo los conceptos matemáticos se eligen y aplican en la formulación de teorías físicas, considerando el problema de la unicidad de las teorías físicas y la posibilidad de unificar teorías divergentes. En este contexto, las aportaciones contemporáneas de Mark Steiner y Sorin Bangu son fundamentales para entender el enigma planteado por Wigner. Steiner destaca las dos posibles interpretaciones del problema de Wigner, centrándose en la relevancia del origen estético de las matemáticas. Por su parte, Bangu ofrece una perspectiva tomando como base la exposición de Steiner, pero con modificaciones para evitar las críticas sobre la consideración de una efectividad significativa de las matemáticas en la física. Finalmente, se presenta un ejemplo concreto del uso de conceptos matemáticos en la física, específicamente en el movimiento armónico simple, para ilustrar cómo conceptos abstractos pueden tener aplicaciones prácticas significativas en la ciencia.

## 2. La misteriosa efectividad de las matemáticas

Wigner comienza su artículo reflexionando sobre la sorprendente adecuación de las matemáticas en la descripción de las leyes de la naturaleza. Para ilustrar este punto, relata la historia de dos amigos, uno estadístico y el otro escéptico, que discuten sobre la utilidad de las matemáticas. En el relato, el estadístico muestra a su amigo un trabajo que comienza con la distribución gaussiana. El estadístico explica el significado de los símbolos en su trabajo, incluyendo el símbolo de pi, que representa la razón de la circunferencia a su diámetro. Su amigo incrédulo no puede entender cómo un símbolo matemático aparentemente tan abstracto puede relacionarse con la población real. Este escepticismo destaca el asombro ante la conexión entre conceptos matemáticos abstractos y su aplicación práctica en la descripción de fenómenos naturales. De esta forma, la anécdota ilustra el asombro que puede generar la conexión entre el mundo de las matemáticas y el mundo físico.

Posterior a la anécdota, Wigner plantea los dos puntos de vista fundamentales que va a desarrollar en su artículo, a saber: ¿Cómo es posible que los conceptos matemáticos se revelen en conexiones completamente inesperadas y con frecuencia permitan una descripción precisa de los fenómenos involucrados en tales conexiones? Y en segundo lugar: ¿Son nuestras teorías matemáticas las únicas correctas? Para responder a las interrogantes, Wigner asegura que no existe una explicación racio-

nal para la eficacia de las matemáticas, por lo que todo lo relacionado con la utilidad de estas bordea con lo enigmático.

Es por esto por lo que, para entender mejor este misterio, Wigner ofrece definiciones claras sobre matemáticas y la física. Reconoce el físico húngaro que en un inicio las matemáticas fueron desarrolladas para describir entidades presentadas por el mundo empírico, pero a medida que se expanden las matemáticas y se establecen nuevos conceptos, estos parecen estar cada vez más lejos de las entidades mundanas. En este sentido, entiende las matemáticas como “la ciencia de operaciones expertas con conceptos y reglas inventados precisamente para ese fin” (Wigner 1960, p. 2). Ciertamente, esta definición subraya el carácter abstracto y aparentemente arbitrario de las matemáticas que, no obstante, resulta ser extraordinariamente útil en la ciencia.

Considerando lo anterior, Wigner explica cómo conceptos avanzados como los números complejos y los operadores lineales fueron ideados inicialmente por su elegancia y belleza formal antes de encontrar aplicaciones prácticas en la física. Estos conceptos, desarrollados sin ninguna intención práctica, demostraron ser útiles en la formulación de teorías físicas que describen con precisión fenómenos naturales. Esto subraya la irrazonable eficacia de las matemáticas: una ciencia creada sin consideración práctica resulta ser indispensable para la comprensión de la naturaleza.

Por otro lado, define la labor del físico como aquella que está enfocada en estudiar las leyes de la naturaleza inanimada (Wigner 1960, p. 3). Wigner, al destacar la relación entre la física y las leyes de la naturaleza, señala que todas las leyes son “afirmaciones condicionales que permiten una predicción de algunos sucesos futuros sobre la base del conocimiento presente” (Wigner 1960, p. 3). Aunque cabe destacar que el mismo físico húngaro reconoce que la utilización de las leyes de la naturaleza para predecir acontecimientos solamente es posible cuando se tiene conocimiento de los factores relevantes del estado presente. Por último, queda recalcar que la argumentación anterior tiene como objetivo principal indicar que las leyes de la naturaleza son siempre afirmaciones condicionales y se refieren únicamente a una fracción muy pequeña de nuestro conocimiento del mundo.

En resumidas cuentas, Wigner realiza una distinción importante entre las matemáticas como un producto de la mente humana y la física como una descripción de las leyes de la naturaleza.

### 3. La relación entre matemáticas y física

Una vez esclarecida la visión sobre la matemática y la física, uno de los puntos clave en el artículo de Wigner es indagar en la relación entre ambas. Wigner detalla cómo las matemáticas se incorporan a las teorías físicas, permitiendo describir fenómenos naturales con una precisión sorprendente. Con este fin, el físico húngaro señala que las matemáticas son utilizadas en la física para evaluar los resultados que se obtienen de las leyes de la naturaleza para “aplicar las afirmaciones condicionales a las condiciones particulares que resultan prevalecer” (Wigner 1960, p. 4). Para esto, es necesario que las leyes de la naturaleza estén formuladas en un lenguaje matemático.

Además, el científico húngaro señala que la física selecciona ciertos conceptos matemáticos para formular las leyes de la naturaleza y que solo utiliza una pequeña cantidad de todos los conceptos matemáticos disponibles. Estos conceptos no son elegidos arbitrariamente, sino que en muchos casos han sido desarrollados de manera independiente por los físicos y luego reconocidos como preexistentes en el ámbito matemático. Incluso la idea de que los conceptos matemáticos se eligen por su simplicidad conceptual es errónea; más bien se eligen por su capacidad para permitir manipulaciones inteligentes y razonamientos brillantes (Wigner 1960, p. 5).

Un ejemplo destacado es el uso del espacio de Hilbert complejo con un producto escalar hermitiano en la mecánica cuántica. Para muchos, los números complejos no son naturales ni sencillos y no surgen de las observaciones físicas. Sin embargo, su uso en la formulación de las leyes cuánticas es casi una necesidad, no solo un truco matemático.

Con estas consideraciones, Wigner señala que una posible explicación del uso de las matemáticas por parte del físico para formular las leyes de la naturaleza es que, en cierto sentido, actúa de manera irresponsable. Esto ocurre cuando encuentra una conexión entre dos cantidades que se asemeja a una conexión matemática conocida y concluye que la conexión es la misma tratada en las matemáticas simplemente porque no conoce otra similar (Wigner 1960, p. 6). Sea cierta o no esta acusación, lo importante es señalar que la formulación matemática de las experiencias físicas a menudo conduce a descripciones precisas de una gran variedad de fenómenos.

Para ilustrar lo anterior, Wigner remite a tres ejemplos específicos: el movimiento planetario, la mecánica cuántica elemental y, por último, la electrodinámica cuántica. Estos tres ejemplos anteriores, que

podrían multiplicarse, ilustran la idoneidad y precisión de la formulación matemática de las leyes de la naturaleza, aunque el físico húngaro reconoce su limitado alcance.

En definitiva, esta sección ha expuesto la posición de Wigner sobre cómo las matemáticas, a través de conceptos como los números complejos y el espacio de Hilbert, permiten una descripción precisa de fenómenos físicos. Este uso selectivo y eficaz de herramientas matemáticas destaca su papel esencial en la formulación de teorías físicas. Sin embargo, para el físico húngaro, este enfoque plantea la cuestión de la unicidad de las teorías físicas y la posibilidad de unificar las diferentes formulaciones existentes. Esta interrogante, que explora la convergencia o divergencia de las teorías en una sola estructura coherente, será expuesta a continuación.

#### **4. El problema de la unicidad de las teorías físicas**

Otro tema importante que Wigner aborda es la unicidad de las teorías físicas. La utilidad de los conceptos matemáticos en múltiples teorías plantea la pregunta de si estas teorías son las únicas posibles o si existen otras formulaciones igualmente válidas. La cuestión que se presenta es si las diversas regularidades, esto es, las diversas leyes de la naturaleza que serán descubiertas, se fusionarán en una única unidad consistente o al menos se aproximarán de modo asintótico a una fusión de ese tipo (Wigner 1960, p. 9).

Para completar lo anterior, el autor presenta un ejemplo ilustrativo mediante la comparación de dos teorías físicas fundamentales: la teoría cuántica y la teoría de la relatividad. Estas teorías, que operan en escalas diferentes (macroscópica y microscópica), utilizan conceptos matemáticos distintos. A pesar de sus diferencias, los físicos creen en la posibilidad de unificar estas teorías en el futuro, aunque reconocen la dificultad actual para lograr esta unificación.

En este contexto, la reflexión final de Wigner encapsula la esencia del problema planteado. El autor considera que:

El milagro de la idoneidad del lenguaje de las matemáticas para la formulación de las leyes de la física es un regalo maravilloso que ni comprendemos ni merecemos. Deberíamos estar agradecidos por ello y esperar que siga siendo válido en la investigación futura y que se extienda, para bien o para mal, para nuestro placer o incluso para nuestra confusión, a ramas más amplias del saber. (Wigner 1960, p. 9)

De este modo, Wigner no solo destaca la extraordinaria capacidad de las matemáticas para describir la física, sino también la incertidumbre y el asombro que acompañan a la búsqueda de una teoría unificada que pueda integrar todas las leyes conocidas de la naturaleza. Este cierre subraya la idea de que, mientras se sigue explorando y desarrollando teorías físicas, el papel de las matemáticas como lenguaje fundamental para la ciencia continúa siendo una fuente de cuestionamiento y reflexión profunda.

## 5. Aclaraciones contemporáneas sobre el problema de Wigner

La exposición de Wigner suscitó en un primer momento la atención mayoritaria de científicos y matemáticos, mientras que el interés filosófico en el tema cambió significativamente con la publicación de los trabajos de Mark Steiner a finales de la década de 1980 y durante la de 1990 (Steiner 1989; 1995; 1998). Steiner fue capaz de precisar las posibles interpretaciones sobre el enigma de Wigner, reconociendo dos versiones del argumento. La primera versión puede ser expresada de la siguiente manera:

Los conceptos  $C_1, C_2, \dots, C_n$  son irrazonablemente efectivos en física, y estos conceptos son matemáticos. Por lo tanto, los conceptos matemáticos (o la matemática, según la exposición de Wigner) son (es) irrazonablemente efectivos en física.

Esta versión es rechazada por el filósofo estadounidense, considerándola un silogismo inválido (Steiner 1998, p. 45). En efecto, para Steiner lo que se puede deducir es algo más limitado, a saber, que solo algunos conceptos matemáticos son irrazonablemente efectivos. Además, esta exposición invita a interrogarse sobre si la efectividad irrazonable de los conceptos matemáticos está ligada al hecho de que los conceptos mismos sean matemáticos.

La segunda versión expuesta por Steiner es notoriamente más completa debido a que toma en consideración las preferencias estéticas de los seres humanos. El argumento puede ser expuesto de la siguiente forma:

Los conceptos matemáticos surgen del impulso estético de los seres humanos. Es irrazonable esperar que lo que surge

del impulso estético en los seres humanos sea significativamente efectivo en la física. No obstante, un número significativo de estos conceptos es considerablemente efectivo en la física. Por lo tanto, los conceptos matemáticos son irrazonablemente efectivos en la física. (Steiner 1998, p. 46)

Esta exposición resalta el carácter matemático del fenómeno en cuestión y aclara que lo irrazonable de la relación entre las matemáticas y la física radica precisamente en el hecho de que los conceptos matemáticos son formulados a partir de la belleza formal que los seres humanos encuentran en ellos. Debido a este origen estético, no tendría sentido asumir que estos conceptos tendrían una gran aplicación o utilidad en el estudio de las leyes y fenómenos naturales.

No obstante, esta formulación puede ser puesta en duda en cuanto a la definición de lo que constituye un número significativo o cuándo se puede considerar que la efectividad es realmente significativa. En este contexto, Sorin Bangu (2016), en su análisis, aborda precisamente las problemáticas que emergen de las críticas al argumento de Steiner, ofreciendo una nueva formulación del argumento de Wigner tomando como punto de inicio lo formulado por Steiner. Bangu formula su argumento de la siguiente manera:

Los conceptos matemáticos modernos se originan en las preferencias estéticas de los matemáticos, es decir, en lo que consideran bello o elegante desde un punto de vista abstracto. En este punto, Bangu destaca que la génesis de los conceptos matemáticos no se basa en necesidades prácticas, sino en criterios estéticos subjetivos. Los matemáticos desarrollan y eligen conceptos basándose en su apreciación de la elegancia y la simplicidad, lo que significa que estos conceptos están diseñados para cumplir con ideales estéticos más que con requisitos funcionales inmediatos.

Esta génesis en un dominio estético subjetivo lleva a la pregunta de si estos conceptos nacidos en un contexto no empírico pueden ser efectivos en el ámbito de la física. La cuestión central es si es razonable esperar que conceptos matemáticos formulados en un contexto puramente estético sean útiles en la física. Dado que la física requiere descripciones precisas y verificables de los fenómenos, la aplicabilidad de conceptos matemáticos estéticamente motivados parece en principio poco probable.

Sin embargo, en la práctica científica se vislumbra que varios conceptos matemáticos formulados en un tiempo anterior han sido inesperadamente útiles en el desarrollo de teorías físicas posteriores. A pesar de

la improbabilidad inicial, la experiencia científica muestra que conceptos matemáticos desarrollados en el pasado han demostrado ser útiles para nuevas teorías físicas. Este hallazgo revela que, en la práctica, la eficacia de estos conceptos no está limitada por su origen estético, sino que pueden jugar un papel crucial en la formulación de teorías físicas que no se había previsto inicialmente.

Por lo tanto, no es razonable que los conceptos matemáticos modernos desarrollados con anterioridad a la formulación de las teorías físicas sean efectivos. Bangu concluye que la sorprendente utilidad de conceptos matemáticos antiguos en nuevas teorías físicas cuestiona la idea de que su origen estético limita su aplicabilidad científica. Esta conclusión desafía la suposición de que la matemática debe estar alineada con el contexto en el que se originó para ser efectiva en la física, revelando una relación más compleja entre las matemáticas y la física (Bangu 2016, p. 14).

Dado que en esta formulación la validez del argumento ya no presenta problemas, las objeciones deben centrarse en la veracidad de las premisas. En efecto, las soluciones propuestas hasta la fecha para el problema de Wigner se han formulado principalmente como críticas a una o más de estas premisas (Bangu 2012, 2016; Bangu y Moir 2018).

En resumen, la evolución del debate sobre la aplicabilidad de las matemáticas ha pasado de la formulación inicial de Wigner a las aclaraciones más recientes de Steiner y Bangu. Mientras que Wigner planteó un problema fundamental, Steiner clarificó y precisó la discusión al considerar como un elemento fundamental en la argumentación de Wigner la influencia de las preferencias estéticas en el desarrollo de conceptos matemáticos, cuestionando así la relación entre la estética y la eficacia matemática. La formulación revisada de Steiner subraya la problemática de que conceptos matemáticos nacidos de criterios estéticos pueden resultar efectivos en la física.

La revisión de Bangu al argumento de Steiner busca evitar el problema de enfrentarse a la cuestión de considerar la definición de lo que constituye un número significativo de casos exitosos o cuándo se puede considerar que la efectividad es realmente significativa. En este sentido, con la validez del argumento de Bangu establecida, la atención debe centrarse en examinar ejemplos específicos que demuestren cómo los conceptos matemáticos formulados en contextos estéticos han mostrado una aplicabilidad significativa en la práctica científica. La próxima sección se dedicará a proporcionar un ejemplo sobre esto, ilustrando cómo los números complejos, a pesar de su origen alejado de

los fenómenos empíricos, contribuyen a la comprensión de fenómenos físicos.

## 6. Aplicación de conceptos matemáticos en la física: el caso del oscilador armónico simple

Sorin Bangu (2012, p. 141) analiza el uso de los números complejos considerando la argumentación de Wigner utilizando como ejemplo el oscilador armónico simple que se describe a continuación: Si se tiene una masa  $m$  unida a un resorte horizontal, la fuerza ejercida es proporcional al desplazamiento desde el punto de equilibrio, conforme a la ley de Hooke. Además, de acuerdo con la segunda ley de Newton, esta fuerza es igual al producto de la masa por la aceleración, siendo la aceleración la segunda derivada de la posición respecto al tiempo. Al igualar estas dos expresiones se obtiene la siguiente ecuación diferencial:

$$m \frac{d^2 x}{dt^2} + kx = 0 \quad (1)$$

Esta expresión matemática representa una ecuación diferencial de segundo orden homogénea que puede resolverse mediante una ecuación característica que toma la forma:

$$Pr^2 + Qr + R = 0 \quad (2)$$

Al ser una ecuación característica cuadrática, sus soluciones pueden ser complejas, de la forma  $z = \alpha + \beta i$ . En este caso, la solución general de la ecuación es:

$$y(x) = e^{\alpha x} [c_1 + c_2 \cos(\beta x) + i(c_1 - c_2) \sin(\beta x)] \quad (3)$$

Es en este punto donde se evidencia la participación de los números complejos en la aplicación de las matemáticas a fenómenos físicos como el movimiento armónico simple. En efecto, es aquí donde se puede entrar a la discusión sobre cómo los números complejos y las ecuaciones diferenciales son utilizados en la ciencia a pesar de que inicialmente parecen abstractos y alejados de la realidad física. Bangu (2012, p. 142) argumenta que, aunque los números complejos y las ecuaciones diferenciales se desarrollaron en matemáticas puramente abstractas, su aplicación en la ciencia es efectiva. Específicamente, se menciona que estas ecuaciones son cruciales en diversas áreas de la física, como la mecánica cuántica y la teoría electromagnética, donde las soluciones



complejas permiten una descripción adecuada de los fenómenos. Estos conceptos matemáticos, a pesar de la falta de una interpretación física obvia, resultan ser increíblemente útiles para describir fenómenos físicos, lo que refuerza la idea de la sorprendente aplicabilidad de las matemáticas en las ciencias físicas.

En resumen, este tema resalta cómo los números complejos, aunque abstractos, son herramientas importantes en la formulación de teorías científicas y en la modelización de fenómenos naturales que de otra manera serían difíciles de comprender o describir con solo números reales. El caso de la discusión del oscilador armónico simple y la aplicación de números complejos ilustran cómo conceptos que parecen abstractos pueden desempeñar un papel fundamental en la descripción y comprensión de fenómenos físicos. Este análisis no solo subraya lo llamativo de la aplicabilidad de las matemáticas que en su momento llamó la atención de Wigner, sino que también reafirma la relevancia de los números complejos en la formulación de teorías científicas.

## 7. Conclusión

En su artículo, Eugene Wigner explora el enigma de por qué las matemáticas, siendo una creación abstracta del intelecto humano, resultan extraordinariamente efectivas en la descripción y predicción de fenómenos naturales. A lo largo de su análisis, Wigner destaca cómo las matemáticas, concebidas inicialmente como herramientas para manipular conceptos abstractos, se han convertido en el lenguaje fundamental mediante el cual se formulan y se comprenden las leyes de la física.

Según Wigner, la física y las matemáticas están tan entrelazadas que las matemáticas ofrecen el marco que los físicos necesitan para modelar y prever fenómenos naturales con mucha precisión. Esta conexión es tan efectiva que resulta bastante sorprendente, ya que no hay una explicación clara de por qué funciona tan bien, lo cual añade un toque de misterio y asombro a cómo interactúan estos dos campos del conocimiento.

Además, la discusión contemporánea sobre el problema de Wigner, enriquecida por las aportaciones de filósofos como Mark Steiner y Sorin Bangu, ha llevado a una comprensión más matizada del enigma. Steiner, al examinar el papel de las preferencias estéticas en el desarrollo de conceptos matemáticos, y Bangu, al cuestionar la relación entre origen estético y aplicabilidad práctica, han proporcionado nuevas perspectivas sobre cómo y por qué los conceptos matemáticos pueden resultar efectivos en la física. Ambas perspectivas ofrecen aclaraciones sobre la

argumentación de Wigner y contribuyen a evitar malentendidos en la búsqueda de por qué los conceptos matemáticos pueden ser sorprendentemente efectivos en la física.

En última instancia, el trabajo de Wigner invita a una reflexión profunda sobre el papel de las matemáticas en la ciencia y su capacidad para capturar y modelar el ámbito físico. El artículo de Eugene Wigner plantea preguntas fundamentales sobre la relación entre las matemáticas y los fenómenos físicos, sugiriendo una conexión profunda que aún no comprendemos completamente. Las implicaciones filosóficas y científicas de este misterio son importantes, desafiando nuestra comprensión de la naturaleza de las teorías científicas y la estructura del universo. En última instancia, el trabajo de Wigner insta a reflexionar sobre el papel de las matemáticas en la ciencia y a continuar explorando el enigma de su sorprendente eficacia.

## Referencias

- Bangu, S. (2012). *The Applicability of Mathematics in Science: Indispensability and Ontology*. London: Palgrave Macmillan.
- Bangu, S. (2016). On ‘The Unreasonable Effectiveness of Mathematics in the Natural Sciences’. En E. Ippoliti, F. Sterpetti, & T. Nickles (eds.), *Models and Inferences in Science* (pp. 11-29). Cham: Springer.
- Bangu, S., & Moir, R. (2018). The ‘Miracle’ of Applicability? The Curious Case of the Simple Harmonic Oscillator. *Foundations of Physics*, 48(9) , pp. 507-525. <https://doi.org/10.1007/s10701-018-0152-5>.
- Frigg, R., & Nguyen, J. (2017). Models and representation. En L. Magnani, & T. Bertolotti (eds.), *Handbook of Model-Based Science* (pp. 49-102). Dordrecht: Springer.
- Steiner, M. (1989). The Application of Mathematics to Natural Science. *The Journal of Philosophy*, 86(9) , pp. 449-480. <https://doi.org/10.2307/2026759>.
- Steiner, M. (1995). The Applicabilities of Mathematics. *Philosophia Mathematica*, 3(2), pp. 129–156. <https://doi.org/10.1093/philmat/3.2.129>.
- Steiner, M. (1998). *The Applicability of Mathematics as a Philosophical Problem*. Cambridge: Cambridge University Press.

Wigner, E. (1960). The Unreasonable Effectiveness of Mathematics in the Natural Sciences. *Communications in Pure and Applied Mathematics*, 13(1), pp. 1-9. <https://doi.org/10.1002/cpa.3160130102>.



# Para una interpretación lógica de la interrogatividad de las preguntas.

Mario Tapia Ramírez<sup>a</sup>

## Resumen

En el presente trabajo nuestro objetivo principal es dar una definición de interrogatividad que utilice la relación de derivabilidad entre la presuposición y el desideratum de una interrogación. Para esto abordaremos la noción de pregunta, primero repararemos en cierto conjunto de asuntos que se dan al preguntar, luego abordaremos el comentario de Gottlob Frege (1918) sobre las preguntas que expresan proposiciones completas, para luego abordar la interpretación de Jaakko Hintikka (1999). En ésta encontramos que una pregunta puede ser interpretada lógicamente como la relación condicional de un desideratum y una presuposición, tal que ambos expresan lo que se quiere saber y lo que actualmente se sabe para preguntar. De esta forma, mostraremos que podemos definir la condición de ignorancia de una pregunta y la de responsabilidad sobre la base de la derivabilidad de estos dos elementos de una pregunta.

**Palabras clave:** lógica interrogativa, responsabilidad, ignorancia, pragmática, inferencia.

## Abstract

In the present work, our main objective is to provide a definition of interrogativity that utilizes the derivability relation between the presupposition and the desideratum of an interrogation. To achieve this, we will approach the notion of a question. Firstly, we will consider a certain set of issues that arise in questioning. Then, we will delve into

---

<sup>a</sup>Miembro Centro de Estudios en Filosofía, Lógica y Epistemología (CEFILOE), Instituto de Filosofía, Universidad de Valparaíso, Valparaíso, Chile.  
Contacto: mario.tap.ram@gmail.com

Gottlob Frege's commentary (1918) on questions expressing complete propositions, followed by Jaakko Hintikka's interpretation (1999). In Hintikka's interpretation, we find that a question can be logically interpreted as the conditional relation between a desideratum and a presupposition, where both express what is desired to be known and what is currently known in order to ask. In this way, we will show that we can define the conditions of ignorance of a question and its answerability based on the derivability of these two elements of a question.

**Keywords:** interrogative logic, answerability, ignorance, pragmatics, inference.

## 1. Introducción

En el presente trabajo queremos hacer una pequeña exposición sobre la noción de interrogatividad, en la que buscamos definirla sobre la base del comportamiento semántico de las preguntas. Para esto tomaremos el trabajo de Gottlob Frege<sup>1</sup> y de Jaakko Hintikka<sup>2</sup> como dos hitos en la interpretación lógica de enunciados interrogativos. El primero por reconocer que hay cierto tipo de proposiciones que tienen un carácter interrogativo y exponer cuáles son los límites de la expresión lógica de un pensamiento interrogativo. El segundo porque es el despliegue teórico que nos permite dar con una forma para las interrogaciones en un contexto en el que son tenidas por objetos epistémicos. Tomando estos dos antecedentes principales mostraremos que el carácter interrogativo de una pregunta puede ser definido por un contexto en el que el agente sepa que hay alguna respuesta y no sepa cuál es la respuesta.

## 2. Sobre preguntar

Para preguntarnos por la interrogatividad necesitamos escudriñar el objeto que privilegiadamente se encuentra en una situación interrogativa: la pregunta. Aquí le llamaremos pregunta al enunciado que en

---

<sup>1</sup>Particularmente en *El Pensamiento. Una investigación lógica* de 1918.

<sup>2</sup>Su trabajo en torno a interrogaciones en *Inquiry as Inquiry. A Logic of Scientific Discovery*, (Hintikka 1999), en especial los trabajos: *Interrogative Logic as a General Theory of Reasoning* (Hintikka, Halonen y Mutanen 2002, que en el momento de la publicación de *Inquiry as Inquiry* estaba en prensa), *A Spectrum of Logic of Questioning* (Hintikka 1985) y *Semantics and Pragmatics for Why-Questions* (Hintikka y Halonen 1995). A este conjunto de textos, en adelante, le llamaremos la teoría interrogativa de Hintikka.

castellano inicia y termina con signos de interrogación. Pese a esta caracterización no serán de nuestra consideración todas las preguntas que no tengan un sentido interrogativo (esperamos esta expresión se entienda claramente al final de nuestra exposición). No consideraremos, por ejemplo, las preguntas retóricas o preguntas proferidas con finalidades muy distantes del sentido habitual de las expresiones en uso<sup>3</sup>. Como primera parte de nuestra aproximación consideraremos la actividad de preguntar, considerando qué hacemos al preguntar, cómo se caracteriza ese preguntar y cuáles son los elementos que participan de esto.

Cuando preguntamos, a alguien le preguntamos por algo. Tanto suscitamos en un agente una acción de respuesta, como hacemos presente la falta de la respuesta, que es aquello por lo que se pregunta. Algo que no hay nos mueve a preguntar. El problema de la comprensión lógica de la interrogación radica en estos dos asuntos: que, por una parte, dada la acción suscitada, requerimos que el lenguaje en uso sea dinámico (o que al menos sea considerado respecto de una dimensión pragmática) y, por otro, dado que es necesario hablar de lo que no hay, necesitamos un lenguaje irrealista.

Una pregunta siempre pregunta por algo y a esto, de modo general, le llamamos respuesta. La respuesta, sea cual fuere la pregunta que la suscita, nos permite inferir una proposición completa. Esto no quiere decir que una pregunta exitosamente respondida permita siempre hacer luego una aserción. Pese a la complejidad de su interpretación lógica, las interrogaciones son objetos lingüísticos rudimentarios y su resultado puede ser variado: por ejemplo, la respuesta exitosa a una pregunta puede ser otra pregunta<sup>4</sup>. Lo que las hace rudimentarias es que disponen de muy pocos recursos, sólo los que tenemos para anunciar una falta: que no hay algo ahí donde debería haberlo; que hay una interrogación donde debería haber una respuesta. Esto quiere decir, también, que al preguntar por algo (al hacer presente la falta de respuesta) necesariamente tenemos que poner algo que indique esa falta y que abra

---

<sup>3</sup>Wittgenstein (1999) en *Investigaciones Filosóficas* ejemplifica esto con alguien que un día soleado en el exterior anuncia ¿no hace un buen día hoy?, donde el objetivo de la interrogación no es saber si hay o no un buen día, sino iniciar una conversación.

<sup>4</sup>Para esbozar este tema puede verse la discusión la respuesta que le da Levi (2003, p. 353) a Olsson, a propósito de su artículo *Possible Answers -to What Question?* (Olsson 2003). Este punto es especialmente interesante, porque si tenemos que el resultado de interrogaciones pueden ser otras, también tenemos que hay expresiones de conocimiento que tienen un cierto carácter interrogativo.

la posibilidad lógica a que eso sea traído. Si no se formulan respuestas posibles en una pregunta no es posible que haya algo por lo que se interroge.

*Corolario.* Si se pregunta por algo, entonces se hace explícita la falta de una respuesta como solicitud de una, luego necesariamente se requiere algún conjunto de respuestas posibles para hacer explícita esa falta.

Siendo las respuestas posibles una condición necesaria de preguntar por algo, la posibilidad de esas respuestas está dirigida a un determinado escenario. Pareciera que estamos dispuestos a lidiar con un contenido incompleto cuando hay otro escenario en el que se pueda completar. La respuesta (el objeto por el que se pregunta, lo que no hay) es proyectada en la pregunta al escenario en el se encontraría (la respuesta) y la pregunta podría tenerse por respondida. En este sentido, no hay preguntar sin que sea preguntar por y en eso está contenida la posibilidad de una respuesta.

El preguntar es una actividad en la que se pregunta por algo a alguien, por lo que en la pregunta se reconocen dos comisiones iniciales<sup>5</sup>: que falta la respuesta y que hay una respuesta (en algún escenario). Sin embargo, también pareciera ser que al preguntar por algo inmediatamente aparece la posibilidad de alguna respuesta, de forma que en cuanto preguntamos por algo estamos obligados a tener la expectativa de alguna respuesta (corolario).

### 3. Preguntas como pensamientos completos

Como suele ocurrir en lógica, la interpretación de Frege (1918) abre un primer camino para considerar el estatus lógico de las preguntas. Sin embargo, hay que considerar que lo que le interesa a Frege es otra cosa.

---

<sup>5</sup>En general en el presente artículo ocuparemos la expresión comisión como coextensiva de la expresión *commitment*. En general, incluso en la literatura especializada, esta expresión se traduce por *compromiso* y creemos que ésta parece dar la impresión de que la acción compromisoria es voluntaria. La expresión *commitment* refiere a un compromiso obligatorio al que se está sometido. Por ejemplo, si uso un auto tengo el compromiso de tener bencina. Acá no hay voluntad, si lo segundo no es el caso, si no cuido de tener bencina, el auto sencillamente no funciona. En el ejemplo, de hecho, funciona más apropiadamente comisión: si uso un auto tengo la comisión de tener bencina. Comisión además destaca el carácter pragmático de *commitment*, que puede implicar una serie de acciones más que admitir (incluso obligatoriamente) alguna determinada expresión.



En *El Pensamiento, una investigación lógica* Frege busca identificar qué exactamente es un pensamiento (*Gedanke*), tal que es el asunto con el que trabaja la lógica (lo que hoy por hoy llamamos proposición). Dentro de este asunto Frege llega al problema de si a las preguntas les corresponde un sentido lógico tal que expresan un pensamiento. A este respecto Frege es muy claro:

En una oración con pronombre interrogativo (*Wortfrage*) enunciamos una oración incompleta que debe obtener un sentido verdadero al ser completada de acuerdo a nuestro requerimiento. Este tipo de preguntas queda por lo tanto fuera de nuestra consideración. (Frege 1918, p. 158)

Las preguntas que comienzan con qué, cómo, cuándo, dónde no tendrían sentido lógico porque requieren ser completadas. Esto quiere decir que rechaza las que expresan una parte de un pensamiento, pero también rechaza las que piden un elemento del lenguaje que complete la oración interrogativa.

La situación es diferente en el caso de las oraciones interrogativas completas (*Satzfragen*). Esperamos oír *sí* o *no*. La respuesta *sí* quiere decir lo mismo que una oración asertiva, pues presenta como verdadero el pensamiento que ya está completo en la oración interrogativa. (Frege 1918, p. 158)

Luego es distinto con las preguntas que tienen una oración completa: expresan un pensamiento completo y pueden ser expresadas lógicamente. La forma lógica de la pregunta se deja ver entre las posibles respuestas (*sí* o *no*) y que el resultado del *sí* es la oración interrogativa en forma asertiva. La forma lógica de una pregunta de esta clase es  $(p \vee \neg p)$ . En primera instancia, tenemos que notar que lo solicitado es un valor (un asunto metalingüístico y no del lenguaje). Segundo, que sea cual fuere el caso, necesariamente hay una respuesta, porque no hay otra posibilidad que sea alguna de esas dos la respuesta. Y tercero, que esa respuesta satura el sentido interrogativo de la pregunta, porque no hay posibilidad de que haya alguna otra respuesta.

Hay algo importante tras el comentario de Frege (al que seguramente no le interesaba, como a nosotros, la noción de interrogatividad): un pensamiento (una proposición) no es una aserción; tiene la forma de una oración pero su verdad no se encuentra juzgada. Con esto Frege intenta distinguir la proposición del juicio. Y (en favor de nuestros intereses) comenta:

Sin querer con esto dar una definición, llamo pensamiento (*Gedanke*) a algo por cuya verdad cabe preguntar. (Frege 2017, p. 216)

De esta forma el interrogar tiene cabida en el marco de Frege sólo en cuanto es determinable la verdad de las proposiciones, en el momento que hay entre la proposición y la determinación de su verdad.

El progreso de la ciencia suele ocurrir así: primero se aprehende un pensamiento más o menos como lo podemos tener en una oración interrogativa completa (*Satzfrage*)- y luego, una vez terminadas las investigaciones emprendidas, se reconoce que es verdadero. En la forma de la oración asertiva expresamos el reconocimiento de la verdad [del pensamiento]. (Frege [1918] 2017, p. 160)

En el caso de Frege las preguntas también están dirigidas a su respuesta, a que sea reconocido que tal proposición es verdadera o falsa.

La responsabilidad es relevante porque nos permite interpretar la pregunta como una proposición con la particularidad de ser verdadera sin que podamos dar cuenta de su verdad: alguna de las respuestas posibles es verdadera, pero no sabemos cuál. En Frege el problema de la responsabilidad se soluciona por definición: sólo podemos preguntar si una determinada proposición es verdadera o falsa y a toda proposición le corresponde o ser verdadera o ser falsa (y nada más). De esta forma, sea cual fuere la pregunta, si tiene forma lógica y expresa un pensamiento completo, entonces es respondible. La responsabilidad nos la asegura la validez de la forma de la pregunta ( $p \vee \neg p$ ): porque la expresión es válida, si es verdadera disyunción, entonces alguno de los dos disyuntos necesariamente es verdadero. Sea el que fuere el verdadero, ese es la respuesta. Y, además, sea cual fuere la respuesta, la otra posibilidad es falsa.

#### 4. Preguntas como interrogaciones proposicionales

En Frege estamos entendiendo una pregunta como un pensamiento interrogativo, i.e. una pregunta es el tipo de proposición que tiene un carácter interrogativo expresado por una disyunción, con la que se solicita la verdad o falsedad de un pensamiento; en Hintikka, por contra, entenderemos una pregunta como una interrogación proposicional,

i.e. una estructura producida a base de proposiciones epistémicas que expresa un contenido interrogativamente.

El punto de partida de la teoría de interrogaciones de Hintikka (1999) será una regla pragmática: si preguntamos, entonces tenemos la comisión de que haya una respuesta. Y si tenemos la comisión de que haya alguna respuesta, entonces se sabe que hay alguna posible respuesta.

Siguiendo a Hintikka, con un conjunto de posibles respuestas definiremos una pregunta: para una pregunta  $Q$  tendremos el conjunto de posibles respuestas  $\{q_1, \dots, q_n\}$ . Aquí, tal como propusimos antes, las respuestas posibles son lo solicitado y la pregunta es la solicitud. La pregunta está compuesta de dos partes: lo que sabemos para poder preguntar, la presuposición, y lo que queremos saber, el desideratum. Con un lenguaje epistémico tradicional, para la pregunta  $Q$ , tendremos que la presuposición será que el agente sabe que alguna de las posibles respuestas es la verdadera,

$$K(q_1 \vee \dots \vee q_n)$$

Y el desideratum expresa que el agente sabe alguna de las posibles respuestas en el escenario en el que la pregunta se encuentra respondida,

$$Kq_1 \vee \dots \vee Kq_n$$

Dado los respectivos significados de la presuposición y del desideratum, en una pregunta el desideratum está condicionado a la presuposición, porque sea lo que fuere que se quiere saber, eso está antecedido por lo que actualmente se sabe para querer saberlo.

## 5. Interrogatividad como ignorancia y responsabilidad

La relación entre el desideratum y la presuposición nos puede indicar varias cosas sobre la pregunta. En primer lugar la fórmula de la pregunta, la relación condicional de un desideratum a una presuposición, no es una fórmula válida. Como parecerá intuitivo, el desideratum (lo que quiere saber) escapa del rango epistémico de lo que el agente sabe, dentro de lo que está la presuposición. Esto es relevante porque, tal como menciona Hintikka (1999), si podemos inferir el desideratum desde la presuposición, entonces la pregunta se encuentra respondida.

Y, luego, si no está respondida, entonces no podemos inferir el desideratum desde la presuposición. Esto resulta intuitivo al pensar la operación de respuesta: hay un antecedente verdadero que muestra una de las posibles respuestas, tal que la presuposición es verdadera, luego el desideratum también. Esto quiere decir que desde la presuposición, un agente no puede inferir el desideratum salvo que la pregunta esté respondida, porque fundamentalmente eso es responder la pregunta: contar con los recursos para llegar al desideratum desde la presuposición.

*Definición (Ignorancia).* Un agente formula una expresión bajo comisión de ignorancia, sin saber cuál es la respuesta, si y sólo si el desideratum no puede ser inferido desde la presuposición, luego la pregunta es actualizante y su respuesta llega a actualizar el estado epistémico del interrogador.

En la vida corriente a ese tipo de preguntas les llamamos honestas: una pregunta cuyo objetivo único es llegar a saber. Llegar a saber quiere decir actualizar el estado de conocimiento. Creemos en sentido estricto que esta es la única comisión pragmática que antecedente un interrogatorio, porque sólo podemos producir una prueba cuando se incumple la condición de ignorancia y la presuposición es suficiente para inferir el desideratum; en caso de que la pregunta cumpla la condición (y si el conjunto de posibles respuestas no fuese finito), sólo confiamos en la admisión de tal comisión, pero no podemos mostrarlo.

A diferencia de la comisión de ignorancia, la comisión de responsabilidad de una pregunta es demostrable. Si con la presuposición aseguramos la responsabilidad de la expresión, entonces al inferir la presuposición desde el desideratum (lo que queremos saber), estamos asegurando la responsabilidad de la pregunta y se da sólo teniendo el desideratum. En términos pragmáticos, basta con que tengamos la comisión de preguntar por algo (como expresamos en el corolario) para que nos veamos forzados a aceptar necesariamente alguna respuesta posible, dado que la pregunta es responsable.

Inferir la presuposición desde el desideratum supone que lo que se quiere llegar a saber necesariamente es posible para lo que se sabe. Eso también quiere decir que dado lo que se quiere saber, si la fórmula de la presuposición asegura la responsabilidad, entonces la pregunta es responsable. De la misma forma que lo central semánticamente de una pregunta es el desideratum, tenemos, luego, algo de ese desideratum que hace sostenible la expresión como fórmula de un lenguaje. No daremos una descripción detallada de los pasos de la demostración por extensión.

Sea  $i$  el nombre de cada una de las respuestas posibles de  $Q$  (tal que  $i = 1, \dots, n$ ):

	3) $[p_i]^{(4)}$	$I\vee$
	4) $p_1 \vee \dots \vee p_n$	$I\rightarrow$
	5) $p_i \rightarrow p_1 \vee \dots \vee p_n$	$I\Box$
	6) $K(p_i \rightarrow p_1 \vee \dots \vee p_n)$	$Ax.K$
	7) $Kp_i \rightarrow K(p_1 \vee \dots \vee p_n)$	$E\rightarrow$
2) $[Kp_i]^{(8)}$	8) $K(p_1 \vee \dots \vee p_n)$	$E\vee$
1) $(Kp_1 \vee \dots \vee Kp_n)$	9) $Kp_i \rightarrow K(p_1 \vee \dots \vee p_n)$	$E\vee$
	10) $K(p_1 \vee \dots \vee p_n)$	

Cuadro 3: Demostración por método de Gentzen de la validez de la inferencia de la presuposición desde el desideratum.

Sin embargo, nos interesa destacar que lo que posibilita la inferencia es el paso 5 (basado en los pasos 3 y 4): en el paso 5 clausuramos la suposición de 3 para la introducción del condicional y construimos la relación entre cada posible respuesta y la parte proposicional (no epistémica) de la presuposición, desde lo que podemos llegar a inferir con 7 y la suposición del caso en cuestión (2), la presuposición, lo que nos permite llegar a 9, lo que aplicado a cada posible respuesta nos permite eliminar la disyunción inicial (1), concluyendo válidamente la presuposición. Por una parte esto muestra cómo es posible construir la inferencia desde el desideratum a la presuposición, pero también muestra que no es posible a la inversa, porque el paso 3 no lo podemos reproducir (salvo que la pregunta se encuentre respondida).

*Definición (Respondibilidad).* Un agente formula una expresión interrogativa bajo comisión de responsabilidad, sabiendo que hay alguna respuesta, si y sólo si la presuposición puede ser inferida desde el desideratum.

Haciendo el recorrido desde lo que podemos interpretar como preguntas en Frege hasta Hintikka, podemos construir lógicamente el esquema intuitivo con el que partimos (al inicio de la primera sección): tenemos un esquema en el que un interrogador representado por un agente epistémico hace una solicitud profiriendo una pregunta en la que se propone un algo que se solicita a alguien. Esa solicitud es acá propuesta como la diferencia entre lo que sabe el agente (la presuposición) y lo que quiere saber (el desideratum). Y la diferencia que nos asegura la interrogatividad en la pregunta está en lo que hemos definido formalmente como dos propiedades de las expresiones interrogativas

(vía escudriñar en las preguntas), a saber la ignorancia y la responsabilidad. Podemos definir interrogatividad de esta forma.

*Definición (Interrogatividad).* Un agente interrogativo formula una expresión bajo comisión de interrogatividad si y sólo la expresión es proferida bajo una comisión de responsabilidad y de ignorancia, esto es, si la presuposición puede inferirse del desideratum, pero el desideratum no puede ser inferido de la presuposición.

La interrogatividad de una expresión la definimos acá por la diferencia semántica de una presuposición y un desideratum en el que el contenido del primero es insuficiente para dar cuenta del segundo, luego esto último supone que la pregunta no se encuentra respondida y, sin embargo, que se puede responder.

De esta forma cuando decimos que una pregunta tiene sentido interrogativo, lo que hacemos es indicar que no se conoce la respuesta y se puede responder, lo que formalmente hemos mostrado por la relación inferencial entre la presuposición y el desideratum. Luego cuando decimos que una expresión (en general) es interrogativa significa que la podemos describir por una relación condicional, donde lo que queremos saber en ella es de mayor alcance epistémico (requiere más recursos) que lo que sabemos de ella.

## 6. Conclusiones

En el presente trabajo hemos hecho una exposición de la noción de interrogatividad explorada a través de la interpretación lógica de preguntas de Gottlob Frege y Jaakko Hintikka.

En primera instancia hemos hecho una aproximación intuitiva a interrogar como acción de la vida ordinaria, para escudriñar cómo estas tienen un rol privilegiado en la noción de interrogación. En esta parte expusimos que el preguntar está compuesto de tres elementos que le dan forma pragmática a la pregunta: alguien que pregunta, algo por lo que se pregunta y alguien al que se le pregunta. Como aquello por lo que se pregunta es algo que falta (la respuesta falta), y esto significa que no hay nada donde debería haber algo, entonces no podemos hacer presente esa falta sino haciendo explícito lo que posiblemente debería haber. Esto es: en una vía de análisis pragmático, no es posible preguntar por algo sin que haya alguna posible respuesta para aquello por lo que se pregunta (corolario).

Luego de esto abordamos las construcciones lógicas para preguntas de Frege y Hintikka. Con Frege pudimos hacer una primera explora-

ción en la que podemos observar que las preguntas proposicionales, que preguntan por la verdad o falsedad de la proposición que enuncian, son las únicas que podrían tener una representación lógica en una perspectiva clásica. Éstas, acá, son proposiciones cuya verdad no ha sido determinada, por lo que su carácter semántico puede ser expresado por la fórmula  $(p \vee \neg p)$ . Esto nos permite inferir que, primero, en la pregunta se expresa el conjunto de posibles respuestas. En estas preguntas hay necesariamente una respuesta y expresan el conjunto total de posibles respuestas. Esta expresión tiene la particularidad de que, sea de lo que fuere que se trate, como a toda proposición le corresponde o ser verdadero o ser falso, entonces cada pregunta de esta forma será necesariamente respondible. Luego con Hintikka encontramos, primero, explícitamente algunos asuntos que en Frege están implícitos y son relevantes para la consideración de la interrogatividad. Segundo, con Hintikka pudimos observar que una pregunta puede ser lógicamente representada bajo un compromiso de respondibilidad, con una presuposición y un desideratum (respectivamente, lo que sabemos para preguntar y lo que queremos saber al preguntar). Un punto importante que tomamos de Hintikka son cualidades de las preguntas sobre la base de la interderivabilidad de la presuposición y el desideratum. Sobre esto destacamos dos puntos: uno, si el desideratum se infiere de la presuposición, entonces la pregunta está respondida; dos, si la presuposición se infiere del desideratum la pregunta es respondible.

Si podemos interpretar la interrogatividad como la capacidad que tiene una pregunta para hacer explícito su carácter semántico, y esto quiere decir que la expresión haga una solicitud realizable, tal que falta la respuesta y hay alguna respuesta posible, entonces se puede definir la interrogatividad como la comisión con la cual una expresión es construida bajo una comisión de ignorancia y de respondibilidad.

## Referencias

- Enqvist, S. (2012). Interrogative belief revision based on epistemic strategies. *Studia Logica*, 100 (3) , pp. 453-479.
- Frege, G. (1918). El Pensamiento. Una Investigación lógica. En A. G.-L. Luis Placencia Ruth Espinoza (Ed.), *Escritos Lógico-Filosóficos* (2017). Argentina, Colihue.
- Hintikka, J. (1999). *Inquiry as Inquiry: A Logic of Scientific Discovery*. Países Bajos, Kluwer Academic Publisher.

Levi, I. (2006). Replies. En E. Olsson (Ed.), *Knowledge and Inquiry. Essays on the pragmatims of Isaac Levi* (pp. 157-166). Estados Unidos, Cambridge University Press.

Olsson, E. (2006). Potential Answers – To What Question? En E. Olsson (Ed.), *Knowledge and Inquiry. Essays on the pragmatims of Isaac Levi* (pp. 157-166). Estados Unidos, Cambridge University Press.

Olsson, E. & Westlund, D. (2006). On the role of the research agenda in epistemic change. *Erkenntnis*, 65 (2) , pp. 165-183.





# S | T

La Serie Selección de Textos es una producción editorial del Instituto de Filosofía de la Universidad de Valparaíso, Chile. Nace en el año 2013 con el propósito de abrir un espacio a los autores para la publicación de libros y capítulos en el área de la filosofía y disciplinas afines. Todos los trabajos son sometidos a arbitraje a doble ciego (double blind review). Sus objetivos generales son: ofrecer publicaciones académicas de calidad científica; proporcionar a la comunidad de académicos y estudiantes un medio de publicación sin fines de lucro; y publicar libros que sean accesibles para todos, sin un costo asociado.

ISBN: 978-956-420-357-7



9 789564 203577