THEORIA

Vol. 41/1 • January 2026

Established in 1952 by Miguel Sánchez-Mazas
Second Series

An International Journal for Theory, History and Foundations of Science

Ὁ Θεὸς ἀριθμετίζει

CALIJ
Centro de Análisis, Lógica e Informática Jurídica (CALIJ)

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

https://ojs.ehu.eus/index.php/THEORIA

T H E O R I A

REVISTA DE TEORÍA, HISTORIA Y FUNDAMENTOS DE LA CIENCIA

AN INTERNATIONAL JOURNAL FOR THEORY, HISTORY AND FOUNDATIONS OF SCIENCE

ESTABLISHED in 1952 by MIGUEL SÁNCHEZ-MAZAS Second Series

THEORIA is a four-monthly journal (issues in January, May and September). Accepted articles are published in Advanced Access as soon as they are ready, prior to their assignment to a specific issue. The contents of this journal are covered, among others, by the following abstracting and indexing services: Arts & Humanities Citation Index®, ISI Alerting Services, Current Contents® / Arts & Humanities, Bulletin Signalétique 519, DICE, ICYT, ISOC and RESH of the Spanish research center CSIC, Mathematical Reviews, Current Mathematical Publications, MathSci, Philosopher's Index, Repertoire bibliographique de la Philosophie, and SCOPUS.

THEORIA's website

http://www.ehu.eus/theoria

Contents

Volume 41/1, January 2026, pp. 1-109

ISSN 0495-4548
e-ISSN 2171-679X

Articles

Articles

Theoria, 2026, 41(1), 5-21
https://doi.org/10.1387/theoria.26626

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

How we learned to stop worrying and love tonk

(Cómo aprendimos a dejar de preocuparnos y a amar a tonk)

Luis Estrada-González
Institute for Philosophical Research, UNAM

Christian Romero-Rodríguez*
Philosophy Department - Universidad del Valle
https://orcid.org/0000-0002-7211-3392

Keywords

Tonk
Dunn semantics
Non-Tarskian logical consequence
Non-transitivity

ABSTRACT: According to common wisdom, the connective *tonk* defined by Prior trivializes any theory that contains it. However, it should not be forgotten that whether an argument holds or not depends to a large extent on the underlying notion of logical consequence. Logical consequence is usually assumed to be Tarskian, that is, reflexive, transitive and monotonic. However, Belnap had already conjectured that *tonk* might not be so problematic in a non-transitive logic, which Cook finally proved in 2005. In this paper we improve on Cook's result in two ways: our hypothesis is simpler (namely, we use fewer interpretations than he did and we do not rely on a disjunctive consequence relation) and it is not *ad hoc* (namely, our working consequence relation is not designed merely to avoid triviality under *tonk*).

Palabras clave

Tonk
Semántica de Dunn
Consecuencia lógica no tarskiana
No transitividad

RESUMEN: Según la sabiduría popular, la conectiva *tonk* definida por Prior trivializa cualquier teoría que la incluya. Sin embargo, no hay que olvidar que, si un argumento es lógicamente válido o no, depende en buena medida de la noción de consecuencia lógica subyacente. Casi siempre se asume que la consecuencia lógica es tarskiana, esto es, que es reflexiva, transitiva y monotónica. Sin embargo, Belnap había conjeturado que *tonk* podría no ser tan problemática en una lógica no transitiva, cosa que finalmente probó Cook en 2005. En este artículo mejoramos el resultado de Cook en dos aspectos: nuestra hipótesis es más simple (a saber, usamos menos interpretaciones que él y no usamos una relación de consecuencia lógica disyuntiva en el *definiendum*) y no es *ad hoc* (a saber, nuestra relación de consecuencia lógica no está definida exclusivamente para evitar la trivialidad).

Gako-hitzak

Tonk
Dunn-en semantika
Ondorio logiko ez-tarskiarra
Ez-trantsitibotasuna

LABURPENA: Iritzi orokorraren arabera, Prior-ek definitutako «*tonk*» lokailuak hura barne hartzen duen teoria oro tribializatzen du. Hala ere, ez da ahaztu behar argumentu baten baliozkotasuna azpian duen ondorio logiko kontzeptuaren mende dagoela, hein handi batean. Oro har, onartzen da ondorio logikoa tarskiarra dela, hau da, erreflexiboa, trantsitiboa eta monotonikoa. Hala eta guztiz, Belnap-ek jada aieru zuen «*tonk*» ez zela hain problematikoa izango logika ez-trantsitibo batean, Cook-ek azkenean 2005ean frogatu zuenez. Artikulu honetan Cook-en emaitza hobetzen dugu, bi modutara: gure hipotesia sinpleagoa da (hau da, Cook-ek baino interpretazio gutxiago erabiltzen ditugu eta ez gara ondorio disjuntiboko erlazio batean oinarritzen), eta ez da *ad hoc* (hau da, guk baliatzen dugun ondorio-erlazioa ez dago berariaz diseinatuta *tonk* lokailuak eragiten duen tribialitatea saihesteko).

* ***Correspondence to:*** Christian Romero Rodríguez. Universidad del Valle, Ciudad Universitaria Meléndez, Calle 13 #100-00, Cali, Valle del Cauca, 760032, Colombia – christian.romero@correounivalle.edu.co – https://orcid.org/0000-0002-7211-3392

How to cite: Romero Rodríguez, Christian (2026). «How we learned to stop worrying and love tonk»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 41(1), 5-21. (https://doi.org/10.1387/theoria.26626).

Received: 19 June, 2024; Final version: 08 July, 2025.
ISSN 0495-4548 - eISSN 2171-679X / © 2025 UPV/EHU Press

Introduction

In *The runabout inference ticket*, Prior (1960) introduced the binary connective *tonk*. Prior defines *tonk* by stipulating its introduction and elimination rules. The problem with *tonk* is that its incorporation into certain theories can trivialize them. It is enough to have *tonk* and a transitive logical consequence relation for it to be possible to conclude any formula.

In *Tonk, Plonk and Plink*, Belnap (1962) noted that the introduction and elimination rules of *tonk* could be compatible with other types of logical consequence relations. That is, the trivialising effects attributed to *tonk* should only be restricted to transitive logical consequence relations. Unfortunately, Belnap never provided a non-transitive logic for this purpose. Many years later, in *What's wrong with Tonk(?)*, Cook (2005) introduced a non-transitive logic for *tonk*, which he called *Tonk Logic* (**TL**). Cook takes the language of *First Degree Entailment* (**FDE**) as a starting point, expands it to include *tonk* and defines a new logical consequence relation. In the non-transitive logic presented by Cook, *tonk* is just another connective in the language and does not trivialize.

However, despite Cook's best efforts, his proposal seems contrived for two reasons. The first is that it requires four admissible interpretations to express *tonk*; the second is that the notion of logical consequence used by him is disjunctive and does not seem to be a minimal modification of the Tarskian notion of logical consequence, so it requires too much conceptual motivation. Here we present a proposal that does not have these problems: we show that three admissible interpretations are sufficient to express *tonk*, and we use a notion of logical consequence that is not disjunctive and is more natural than Cook's in the sense that, when used in contexts with only two admissible interpretations, it is coextensive with the Tarskian notion of logical consequence.

The plan for this paper is as follows. In the first section, we introduce *tonk* and the assumptions necessary for its trivialisation. We also present Barceló's (2008) method for converting (bivalent) truth tables into derivation rules (and vice versa). This method is useful for us to present the evaluation conditions of *tonk*. In the second section, we present Cook's **TL** logic and two objections to his proposal. In the third section, we show that there is a way in which *tonk* can validate Prior's rules without appealing to a logical consequence relation, which may seem artificial. Finally, we present a method for identifying what kind of connective *tonk* is, to determine that it is both a conjunction and a disjunction.

1. *The bomb*

Prior (1960) argued that a meaningful connective cannot be specified simply by its introduction and elimination rules. To illustrate his thesis, he proposed a binary connective called "*tonk*", represented here by "$\ddagger$", with the following introduction and elimination rules, where A and B are any formulas of the language:

tonk Introduction (I$\ddagger$)

$A \vdash A \ddagger B$

$B \vdash A \ddagger B$

tonk Elimination (E$\ddagger$)

$A \ddagger B \vdash A$

$A \ddagger B \vdash B$

The problem with *tonk* is that it seems to trivialize any theory **T** in which it appears. Consider the following proof:

1. $A \vdash_{\mathbf{T}} A ⋚ B$ I⋚
2. $A ⋚ B \vdash_{\mathbf{T}} B$ E⋚
3. $A \vdash_{\mathbf{T}} B$ 1, 2 Transitivity of $\vdash_{\mathbf{T}}$

The conclusion expresses that, if a theory **T** includes *tonk*, any formula of the language entails any other.

For many, including Prior himself, the lesson of *tonk*'s case is that for a connective to be meaningful it must correspond to a pre-theoretical meaning which cannot simply be captured in a system of rules. Some, like Stevenson (1961) in *Roundabout the runabout inference-ticket*, argue that such a pre-theoretical meaning can be captured, at least to a satisfactory extent, by evaluation conditions such as those underlying truth tables. If this is so, then to show that there can be no pre-theoretical meaning associated with *tonk*, it would suffice to show that *tonk* cannot be evaluated, that it has conditions that cannot be satisfied. Stevenson shows that *tonk* has classically impossible conditions to satisfy, but here we want to present an alternative argument which is clearer to us.

Axel Barceló (2008) developed a method for converting two-valued truth tables into natural deduction rules (NDR) with multiple-conclusions[1] and vice versa, which is particularly useful for showing the peculiarities of the evaluation conditions required by *tonk*. Let us first note that in a multiple-conclusion framework, both rules for *tonk* can be simplified:

Tonk Introduction (I⋚)	*Tonk* Elimination (E⋚)
$A, B \vdash A ⋚ B$	$A ⋚ B \vdash A, B$

We now explain Barceló's method. Let ©($A_1, \ldots, A_n$) be a formula whose main connective is the n-ary connective ©. Then

I. ©($A_1, \ldots, A_n$) is a conclusion in an NDR if and only if it is true.

II. ©($A_1, \ldots, A_n$) is a premise in an NDR if and only if it is false.

III. A_k is a premise in an NDR if and only if it is true.

IV. A_k is a conclusion in an NDR if and only if it is false.

Where $1 \leq k \leq n$.

For example, consider the truth table of the extensional conditional (→), where truth and falsity are represented by 1 and 0.

A	*B*	*A*→ *B*
1	1	1
1	0	0
0	1	1
0	0	1

[1] For more on multiple-conclusions, see Shoesmith and Smiley (1978).

From each line of the table, we can obtain an NDR. In descending order, the four rules are as follows:

1. $A, B \vdash_L A \to B$
2. $A, A \to B \vdash_L B$
3. $B \vdash_L A \to B, A$
4. $\vdash_L A \to B, A, B$

In the first case, by (III), *A* and *B* are premises because they are both true. By (I), *A*→ *B* is a conclusion because it is true. In the second case, by (III), A is a premise because it is true. By (II), *A*→*B* is a premise because it is false. By (IV), *B* is a conclusion because it is false. The same can be done with 3 and 4.

The evaluation conditions obtained from Prior's rules for *tonk*, using Barceló's method, are as follows:

A⪼*B* is true if and only if *A* is true.

A⪼*B* is true if and only if *B* is true

A⪼*B* is false if and only if *A* is false

A⪼*B* is false if and only if *B* is false

Or, rewriting:

A⪼*B* is true if and only if *A* is true or *B* is true.

A⪼*B* is false if and only if *A* is false or *B* is false

Here we can clearly see that this connective is undefinable in classical logic, since it requires interpretations that are not admissible in a usual two-valued semantics. Suppose *A* is true and *B* is false: according to the evaluation conditions just given, *A*⪼*B* is true (since *A* is true) and false (since *B* is false). The same result is obtained by assuming that *A* is false and *B* is true. The table for *tonk* would look like this:

A⪼*B*	1	0
1	1	1,0
0	1,0	0

One might ask whether this is not a semantics-dependent rather than a logic-dependent result; in other words, one might ask whether *tonk* is not definable in some other semantics for classical logic. The short answer is "no": there is no semantics *S* which has the following properties:

— *S* is functionally complete using a classical metatheory.

— A connective with homophonic evaluation conditions for *tonk* is definable in *S*.

— The Prior rules for *tonk* are valid under *S*.

— The notion of logical validity is Tarskian.

So far, the bomb: *tonk* trivializes any theory with a semantics S that has the conditions just stated.

In *On three-valued presentations of classical logic*, Da Ré, Szmuc, Chemla and Egré (2023) introduced some logics, with three-valued semantics, which could be considered as presentations of classical logic. However, some of these logics are not reflexive or transitive. Moreover, some of these semantics are not functionally complete. Therefore, even if connectives such as *tonk* are definable in some of these semantics (such as the *tonk* we will present in the next section), they do not satisfy the requirement of being Tarskian or of being functionally complete enough to constitute a counterexample to what has been said above.[2]

2. How we stopped worrying

As we have seen, $A \$ B$ is true and false if one of its components is true and the other is false. For *tonk* to be expressible, then, at least three admissible interpretations are needed for any formula A: A is (only) true, A is (only) false, and A is both true and false. Fortunately for us, and for *tonk*, such semantics are better known and more workable than they were in Stevenson's time.

Before presenting our proposal, we present Cook's (2005) proposal. To do this, it is necessary to first introduce Dunn semantics. Dunn semantics allows us to relate, not necessarily in a functional way, propositional variables to only two truth values, namely, truth and falsity. See more in Dunn (1976). Despite the two-valued nature of the semantics, we can have different sets of admissible interpretations to represent different logics. Omori and Sano (2015) proposed a general method to transform truth tables with up to four admissible interpretations into truth and falsity conditions. In the following, we present Cook's Tonk Logic (**TL**), which uses a Dunn two-valued semantics.

Let L be a formal language for **TL** with a set of formulas constructed in the usual way, from a set of propositional variables VAR = $\{p_1,\ldots, p_n\}$ with the connectives $\sim$, $\wedge$, $\vee$ and $\otimes$, in which '$\otimes$' is the symbol Cook uses for *tonk*.[3] We will use the letters 'A', 'B', 'C',..., of the Latin alphabet as arbitrary formulas of L, and the letters 'Γ', 'Δ',..., of the Greek alphabet as sets of formulas.

An interpretation i for L is a relation between atomic formulas and the truth values (1 and 0), such that we have the following ways:[4]

---

2 In *Formalization of Logic*, Carnap (1943) presented a non-normal semantics for classical logic. In these semantics, interpretations are considered non-normal because, according to Church (1943, p. 493), they "contravene in some way the usual interpretation of classical truth tables". Some ways of "contravening" the usual semantics are to increase the number of values or the number of interpretations, or to force non-equivalence between true (respectively false) and non-false (respectively non-true). For example, in Carnap's non-normal semantics, in the evaluation conditions of negation (expressed here with 'N'), it is either the case that, i) NA is true if and only if A is true, or ii) NA is false if and only if A is not false. Tables compatible with these non-normal semantics can also be found in Church (1953). However, these semantics are not functionally complete either, so they do not provide a counterexample to the above. Some criticisms and comments on Carnap's proposal can be found in the aforementioned Church (1944).

3 There is a conditional $A \rightarrow B$ which is definable as $\sim A \vee B$; and a biconditional $A \leftrightarrow B$ which is definable as $(A \rightarrow B) \wedge (B \rightarrow A)$.

4 We have chosen to work with a Dunn-style bivalent semantics, which is characterized by having only two truth values. Unlike other approaches, the evaluation here is not restricted to a function but is a relation in general. This allows for the consideration of all four possible interpretations, even with only two truth values. The interpretation i should not be understood as a function, it is used as relation that assigns sets of values to formulas. So the notation $i(p)=\{1\}$ does not presuppose a standard functional interpretation. We are not assuming a classical set theoretic metatheory. Many three-valued and many-valued logics can be presented using a Dunn-style bivalent semantics. Presenting them in this way has significant advantages, as it avoids engagement in debates about the ontological or semantic status of other kinds of values (such as i, b, n, etc.), and allows us to bypass discussions concerning whether such entities genuinely count as truth values. Since this issue goes beyond the scope of the present article, we refer the reader to Estrada-González (2019) and (2022) for a more detailed treatment.

https://doi.org/10.1387/theoria.26626

— p_i is true but not false, represented by "$1 \in i(p_i)$ and $0 \notin i(p_i)$"; more briefly, $i(p_i)=\{1\}$;

— p_i is true but also false, represented by "$1 \in i(p_i)$ and $0 \in i(p_i)$"; more briefly, $i(p_i)=\{1, 0\}$;

— p_i is neither true nor false, represented by "$1 \notin i(p_i)$ and $0 \notin i(p_i)$"; more briefly, $i(p_i)=\{\ \}$;

— p_i is false but not true, represented by "$0 \in i(p_i)$ and $1 \notin i(p_i)$"; more briefly, $i(p_i)=\{0\}$.

The interpretations extend to evaluations for all formulas according to the following evaluation conditions[5]:

$1 \in i({\sim}A)$ iff $0 \in i(A)$
$0 \in i({\sim}A)$ iff $1 \in i(A)$
$1 \in i(A \wedge B)$ iff $1 \in i(A)$ and $1 \in i(B)$
$0 \in i(A \wedge B)$ iff $0 \in i(A)$ or $0 \in i(B)$
$1 \in i(A \vee B)$ iff $1 \in i(A)$ or $1 \in i(B)$
$0 \in i(A \vee B)$ iff $0 \in i(A)$ and $0 \in i(B)$
$1 \in i(A \otimes B)$ iff $1 \in i(A)$
$0 \in i(A \otimes B)$ iff $0 \in i(B)$

The evaluation conditions can be presented in tabular form as follows:

$\sim A$	A
{0}	{1}
{1,0}	{1,0}
{ }	{ }
{1}	{0}

$A \vee B$	{1}	{1,0}	{ }	{0}
{1}	{1}	{1}	{1}	{1}
{1,0}	{1}	{1,0}	{1}	{1,0}
{ }	{1}	{1}	{ }	{ }
{0}	{1}	{1,0}	{ }	{0}

$A \wedge B$	{1}	{1,0}	{ }	{0}
{1}	{1}	{1,0}	{ }	{0}
{1,0}	{1,0}	{1,0}	{0}	{0}
{ }	{ }	{0}	{ }	{0}
{0}	{0}	{0}	{0}	{0}

$A \otimes B$	{1}	{1,0}	{ }	{0}
{1}	{1}	{1,0}	{1}	{1,0}
{1,0}	{1}	{1,0}	{1}	{1,0}
{ }	{ }	{0}	{ }	{0}
{0}	{ }	{0}	{ }	{0}

The logical consequence relation of **TL** is the following: Let A and Γ be a formula and a set of formulas of L, respectively. A is a logical consequence of Γ in **TL**, $\Gamma \vDash_{TL} A$, if and only if, either, for each interpretation i, if $1 \in i(B)$, for all $B \in \Gamma$, $1 \in i(A)$; or, for each interpretation i, if $0 \in i(A)$, $0 \in i(B)$ for some $B \in \Gamma$. According to **TL**'s definition of logical consequence, an argument is logically valid if and only if either: either truth is preserved from premises to conclusion in every interpretation (truth-preserving), or falsity is preserved from conclusion to premises, also in every interpretation (falsity-preserving).

[5] We use the same notation for interpretations and evaluations.

Transitivity is not valid in **TL**:

If $A \vDash_{\mathrm{L}} B$ and $B \vDash_{\mathrm{L}} C$, then $A \vDash_{\mathrm{L}} C$

To give a counterexample to Transitivity is to show that $A \vDash_{\mathrm{L}} B$ and $B \vDash_{\mathrm{L}} C$ are valid arguments, while $A \vDash_{\mathrm{L}} C$ is an invalid argument. Specifically, using *tonk*, the following argument is invalid:

If $A \vDash_{\mathrm{TL}} A \otimes B$ and $A \otimes B \vDash_{\mathrm{TL}} B$, then $A \vDash_{\mathrm{TL}} B$

Using the *tonk* table, we can see that if A is true, then $A \otimes B$ is true, so $A \vDash_{\mathrm{TL}} A \otimes B$ is a valid argument. It can also be seen that if B is false, then $A \otimes B$ is false, so $A \otimes B \vDash_{\mathrm{TL}} B$ is valid. However, $A \vDash_{\mathrm{TL}} B$ is not valid because it preserves neither the truth of the premises to the conclusion nor the falsity of the conclusion to the premises. For example, if $i(A)=\{1\}$ and $i(B)=\{0\}$. A consequence of **TL** being a non-transitive logic is that Prior's triviality proof does not hold for it. Thus Cook ends his attempt to claim *tonk* as a meaningful connective.

However, and despite Cook's best efforts, his proposal seems contrived for two reasons. The first is that it requires four admissible interpretations to express *tonk*; the second is that the notion of logical consequence used by him is disjunctive. We begin by discussing the second reason. It seems to us that defining a disjunctive logical consequence is not a minimal modification of the Tarskian notion of logical consequence because it requires too much conceptual motivation not to seem *ad hoc*. Let us recall the Tarskian notion of logical consequence:

— An argument $\Gamma \vDash_{\mathrm{L}} A$ is *logically valid* if and only if, for each interpretation i, if $1 \in i(B)$, for all $B \in \Gamma$, $1 \in i(A)$.

The definition of logical consequence of **TL** can be obtained from the Tarskian notion of logical consequence by putting its *definiens* in disjunction with that of the following logical consequence relation:

— An argument $\Gamma \vDash_{\mathrm{L}} A$ is logically valid if and only if, for each interpretation i, if $0 \in \mathrm{i}(A)$, $1 \in \mathrm{i}(B)$ for some $B \in \Gamma$.

Cook does not give an argument in favor of his disjunctive logical consequence relation because he considers that it is not too "extravagant or inconceivable" (Cook, 2005, p. 221). According to Cook, accepting his consequence relation implies that "there is no reason to prefer the Truth-Preserving Consequence to the Falsity-Preserving Consequence, or vice versa" (Cook, 2005, p. 222). Indeed, when classical logic is presented, for example, one can use either the Truth Preservation Consequence or the Falsity Preservation Consequence interchangeably, since both determine the same collections of valid and invalid arguments. From the classical perspective there is thus no reason to prefer one over the other.

There is a reason why we consider the disjunctive logical consequence relation to be an *ad hoc* hypothesis, namely that the preference of the disjunctive logical consequence definition over one that uses only one of the two disjuncts depends entirely on the presence of *tonk* in the language. Say that a disjunctive logical consequence relation **D** is *redundant* if and only if there is a logical consequence relation **N** in whose *definiens* only one of the disjuncts of the *definiens* of **D** is needed, and both **N** and **D** characterize the same arguments (based on a semantics S).

In logics such as **FDE** or classical logic, the disjunctive logical consequence relation characterizes exactly the same arguments as the logical consequence relations based on any of the disjuncts of the disjunctive relation. In **TL**, the logical consequence relation is not redundant, some arguments are valid thanks to one of the disjuncts, but others are valid thanks to the other disjunct (as they would be invalid according to the first one). For example, $A \mathbin{\lessgtr\!\!\!\!\!|\;} B \vdash_{\mathrm{T}} B$ is a valid argument with Falsity Preservation, but invalid with Truth Preservation. However, if *tonk* were not part of the language of **TL**, as in **FDE**, the logical consequence relation would be redundant. So, the basis of the disjunctive definition has no justification independent of the possibility of defining *tonk*.

https://doi.org/10.1387/theoria.26626

As in other fields, *ad hoc* hypotheses in logic are questionable because they are introduced for the sole purpose of resolving a discrepancy between the theory and particular results. *Ad hoc* hypotheses, such as the disjunctive logical consequence relation, do not extend the scope of the theory or its ability to solve other logical problems. Here we argue that three admissible interpretations are sufficient to express *tonk*, and we use a notion of logical consequence which is not disjunctive, and therefore its justification is not based on the fact that it allows *tonk* to be defined.

Let us recall Prior's argument:

1. $A \vdash_T A \nlessgtr B$ I$\nlessgtr$
2. $A \nlessgtr B \vdash_T B$ E$\nlessgtr$
3. $A \vdash_T B$ 1, 2 Transitivity of $\vdash_T$

Prior's rules for *tonk* do not in themselves entail triviality, as Belnap (1962) had already pointed out. The conclusion depends crucially on the validity of Transitivity of $\vdash_T$ and not only on Prior's rules. So it seems that there are at least two ways to avoid the triviality result that every formula entail every other formula, either one of Prior's rules must be invalid, or Transitivity must be invalid.

Let us recall the evaluation conditions obtained by Barceló's method:

$A \nlessgtr B$ is true if and only if A is true or B is true

$A \nlessgtr B$ is false if and only if A is false or B is false

The table for *tonk* would be as follows, using only three admissible interpretations:

$A \nlessgtr B$	{1}	{1, 0}	{0}
{1}	{1}	{1, 0}	{1, 0}
{1,0}	{1, 0}	{1, 0}	{1, 0}
{0}	{1, 0}	{1, 0}	{0}

However, it seems that, in this semantics, the road to non-transitivity is closed, since the notion of Tarskian logical consequence is transitive, and that rather the rule E$\nlessgtr$ is invalid. In fact, consider the interpretation in which A is {1} and B is {0}. In this case, the premise of $A \nlessgtr B \vdash_T B$ is true but the conclusion is not. On the other hand, the way to define Cook's *tonk* using only these three admissible interpretations is also closed: Cook's *tonk* ($\otimes$) is not true and not false even if A is false and B is true. That is, $i(A \otimes B) = \{\,\}$ when $i(A) = \{0\}$ and $i(B) = \{1\}$. This is not the case for $\nlessgtr$. Using the same evaluation conditions of $\nlessgtr$, we can also define *tonk* for semantics with four admissible interpretations:

$A \nlessgtr B$	{1}	{1,0}	{ }	{0}
{1}	{1}	{1,0}	{1}	{1,0}
{1,0}	{1,0}	{1,0}	{1,0}	{1,0}
{ }	{1}	{1,0}	{ }	{0}
{0}	{1,0}	{1,0}	{0}	{0}

A corollary of this is that Dunn semantics allows us to identify at least two *tonk* connectives. The first, ${\not\lessgtr}_p$, is defined by Prior-style introduction and elimination rules; the second, ${\not\lessgtr}_m$, by evaluation conditions.[6] In general, these are different connectives, for the rule E${\not\lessgtr}$ may not hold for ${\not\lessgtr}_m$, while the connective for which both E${\not\lessgtr}$ and I${\not\lessgtr}$ hold has different evaluation conditions than ${\not\lessgtr}_m$.

According to Barceló's method, *tonk* is a connective that has the truth condition of a disjunction and the falsity condition of a conjunction. However, $\otimes$ has neither the truth-condition of a disjunction nor the falsity-condition of a conjunction. $\otimes$ recovers only part of the evaluation conditions of conjunction and disjunction. For this reason, our *tonk* connective (${\not\lessgtr}_m$ according to the notation of the previous paragraph) is not the same connective as Cook's connective ($\otimes$). Cook's $\otimes$ connective is designed on the basis of the minimal requirements that introduction and elimination rules must fulfil in order to trivialize a logic such as the classical one. For this reason, we consider $\otimes$ to be more similar to ${\not\lessgtr}_p$ than to ${\not\lessgtr}_m$.

Something very similar occurs with the *tonk* defined in Ripley (2015). In that work, *tonk* is introduced through sequent rules, without reference to evaluation conditions. Within that approach, sequents are taken to specify aspects of the meaning of *tonk*. The proposed rules are as follows:

$$\textit{tonk}\mathrm{R}\ \frac{\Gamma \vdash A, \Delta}{\Gamma \vdash A\ \textit{tonk}\ B, \Delta} \qquad\qquad \textit{tonk}\mathrm{L}\ \frac{\Gamma, B \vdash \Delta}{\Gamma, A\ \textit{tonk}\ B \vdash \Delta}$$

To illustrate why *tonk* is problematic, Ripley argues that the Cut rule plays a central role in the derivation of triviality. In doing so, she opens the way for investigations such as ours, which not only aim to diagnose the problem, but also to provide a concrete tool capable of accomplishing the work that Ripley had already outlined. On the other hand, using the method proposed by Béziau (2001, p. 374), it is possible to extract the evaluation conditions of Ripley's *tonk* from its associated sequent rules. These conditions match exactly those of the $\otimes$ connective. Ripley explores the possibility of having her sequent rules validate Prior's original principles. This is why her approach resembles Cook's more closely. Therefore, Ripley's *tonk* is more closely related to ${\not\lessgtr}_p$ than to ${\not\lessgtr}_m$. In this sense, our proposal can be seen as extending Ripley's result by exploring alternative ways of modeling a connective like *tonk*.

The question now is whether there is any way in which Prior's rules E${\not\lessgtr}$ and I${\not\lessgtr}$ can hold for ${\not\lessgtr}_m$, that is, whether we can have it all: a *tonk*-like connective for which Prior's rules are valid and which is described model-theoretically by the truth condition of a disjunction and the falsity condition of a conjunction. The answer is yes, and the simplest way to obtain it is by means of a non-transitive notion of logical validity. Unlike Cook and Ripley's proposals, our approach extracts *tonk* directly from evaluation conditions. We have shown that extraction alone does not ensure the validity of Prior's rules; additional steps are necessary. Ripley and Cook begin oppositely, assuming the validity of the rules from the outset and developing formal mechanisms to avoid triviality. Their versions of *tonk* preserve the truth of Prior's rules within non-trivial frameworks, whereas our proposal does not start from this assumption. Through Barceló's method, we show that a non-transitive context remains necessary for these rules to be valid.

3. *How we learned to love tonk*

One problem with classical logic is that it identifies many properties that perhaps should not be identified. This has been repeated many times for the case of contradiction and triviality, for example, as well as for the pairs falsity/non-

6 In a different approach, this distinction between *tonks* was also explored by Buacar (2018) and Teijeiro (2020).

https://doi.org/10.1387/theoria.26626

truth and truth/non-falsity, which is sufficiently clear in the case of Dunn semantics. Precisely the distinction between falsity and non-truth, on the one hand, and between truth and non-falsity, on the other, makes it possible to distinguish in turn between different notions of logical consequence which are equivalent in classical logic:

tt: An argument $\Gamma \vDash_L \Delta$ is logically valid if and only if, for all interpretation, if the premises are not only false, at least one conclusion is not only false either.

ts: An argument $\Gamma \vDash_L \Delta$ is logically valid if and only if, for all interpretation, if the premises are not only false, at least one conclusion is only true.

st: An argument $\Gamma \vDash_L \Delta$ is logically valid if and only if, for all interpretation, if the premises are only true, at least one conclusion is not only false.

ss: An argument $\Gamma \vDash_L \Delta$ is logically valid if and only if, for all interpretation, if the premises are only true, at least one conclusion is also true.

The properties that premises and conclusions must meet in a logically valid argument are called "standards" (see more in Chemla *et al.* 2017, p. 2198). In the logical consequence relations *tt* and *ss*, the standards of premises and conclusions are the same; in *ts* and *st*, they are different, which is why these logical consequence relations are called "mixed". For more on *tt*, *ss*, *ts* and *st*, see Cobreros *et al.* (2012a). For more on mixed logics, see Cobreros *et al.* (2012b). We have provided the definition of these logical consequence relations, and given the framework of Dunn semantics, they are useful for defining a formal consequence relation when working with three or four interpretations. The only variation lies in the interpretations considered (specifically, what is understood as *tolerant*).

For example, consider the admissible interpretations for **K3**, **LP**, and **FDE**. In these cases, the interpretations to be considered for an argument to be *st*-valid are as follows:

— **K3**: {{1}, { }, {0}}
  Standard for premises: {1}
  Standard for conclusion: {1} and { }

— **LP**: {{1}, {1,0}, {0}}
  Standard for premises: {1}
  Standard for conclusion: {1} and {1,0}

— **FDE**: {{1}, {1,0}, { }, {0}}
  Standard for premises: {1}
  Standard for conclusion: {1}, { }, and {1,0}

When working with three values, as in some semantic presentations of **K3** and **LP**, it is common to assume that there are different interpretations of the third value. In **K3**, this value is typically interpreted as "neither true nor false", while in **LP**, it is interpreted as "both true and false". Now, when considering mixed consequence relations (such as the *st* and *ts* relations), premises and conclusions may treat this additional value differently, which might lead one to think that two different readings of the same value are being used. As a result, it might seem that there are not just three, but actually four values in play.

However, this objection, which could be extended to claim that four values or interpretations are actually being used, does not apply in our case. The apparent difference in how the intermediate value is treated between premises and conclusion is not due to multiple readings of that value, but rather to the differentiated use of strict and tolerant standards in mixed consequence relations. In other words, even though we are working with three interpretations, there is no third value in our semantics that admits multiple interpretations. Thus, it is clear that there are semantic presentations of the *st* and *ts* consequence relations that, at a conceptual level, do not require assuming two readings of the intermediate value.[7]

In classical logic, the direction in which the standards are connected is also indistinct, i.e. it does not matter whether the connection of standards goes from premises to conclusions or from conclusions to premises. In the four relationships above, the connection is from premises to conclusions. But in contexts in which a distinction is made either between truth and non-falsity or between falsity and non-truth, it is also important to distinguish the direction of connection of standards (see more in Wansing and Shramko, 2008).[8] Thus, we have the following notions of logical consequence:

$_{\leftarrow}$*ff*: An argument $\Gamma \vDash_{\mathrm{L}} \Delta$ is logically valid if and only if, for all interpretation, if at least one conclusion is only false, the premises are also false.

$_{\leftarrow}$*fnt*: An argument $\Gamma \vDash_{\mathrm{L}} \Delta$ is logically valid if and only if, for all interpretation, if at least one conclusion is not only true, the premises are only false.

$_{\leftarrow}$*ntf*: An argument $\Gamma \vDash_{\mathrm{L}} \Delta$ is logically valid if and only if, for all interpretation, if at least one conclusion is only false, the premises are not only true.

$_{\leftarrow}$*ntnt*: An argument $\Gamma \vDash_{\mathrm{L}} \Delta$ is logically valid if and only if, for all interpretation, if at least one conclusion is not true, the premises are not true either.

Theorem. I$\lessgtr$ and E$\lessgtr$ evaluated with $_{\leftarrow}$*ntf* are valid, but Transitivity is not.

Proof. Suppose that $A \lessgtr B$ is not true, that is, $1 \notin i(A \lessgtr B)$. This implies, in the semantics we are working on, that it is false, i.e., that $0 \in i(A \lessgtr B)$. But if $0 \in i(A \lessgtr B)$ and $1 \notin i(A \lessgtr B)$ then, by the evaluation conditions of $A \lessgtr B$, both $0 \in i(A)$ and $1 \notin i(A)$, and $0 \in i(B)$ and $1 \notin i(B)$. Thus, for any interpretation i, if $1 \notin i(A \lessgtr B)$, $0 \in i(A)$ and $0 \in i(B)$. But this means that $A \vdash_{\mathrm{T}} A \lessgtr B$ and $B \vdash_{\mathrm{T}} A \lessgtr B$ are $_{\leftarrow}$*ntf*-valid.

Suppose now that A is not true, that is, $1 \notin i(A)$. This implies, in the semantics we are working on, that it is false, i.e., that $0 \in i(A)$. Then, by the evaluation conditions of $A \lessgtr B$, $0 \in i(A \lessgtr B)$. Therefore, since i was arbitrary, $A \lessgtr B \vdash_{\mathrm{T}} A$ is $_{\leftarrow}$*ntf*-valid. The reasoning for $A \lessgtr B \vdash_{\mathrm{T}} B$ is the same, *mutatis mutandis*.

Finally, there is an interpretation that shows that $A \vdash_{\mathrm{T}} B$ is not $_{\leftarrow}$*ntf*-valid, namely, $i(A) = \{1\}$ and $i(B) = \{0\}$. Thus, while $A \vdash_{\mathrm{T}} A \lessgtr B$ and $A \lessgtr B \vdash_{\mathrm{T}} B$ are $_{\leftarrow}$*ntf*-valid, $A \vdash_{\mathrm{T}} B$ is not $_{\leftarrow}$*ntf*-valid.

Considering the language of **FDE** and the consequence relation $_{\leftarrow}$*ntf*, we obtain the same valid arguments as those in **ST**.[9] (For more information, see Barrio, Rosenblatt and Tajer (2015).) In the resulting logic, Transitivity does not hold. However, all logical truths of classical logic[10] and classically valid arguments of the form $\Gamma \vdash_{\mathrm{L}} A$ (though not all meta-arguments, i.e., arguments of the form "if $\Gamma \vdash_{\mathrm{L}} A$ then $\Delta \vdash_{\mathrm{L}} B$"; Transitivity is an example of this) do hold.[11] One advantage of adopting $_{\leftarrow}$*ntf* is that it is not disjunctive, so with or without t*onk*, it does not seem *ad hoc*. Supposing the validity

7 For example, in the presentation of the four definitions of mixed logical consequence proposed by Cobreros *et al.* (2012a), more than two truth values are employed. This raises the question, as one of the reviewers suggests, whether three or four distinct values are being used.

8 As a reviewer correctly observed, in classical contexts where Transitivity is a valid meta-argument, the relations *tt*, *ts*, *st*, and *ss* are coextensional with $_{\leftarrow}$*ff*, $_{\leftarrow}$*fnt*, $_{\leftarrow}$*ntf*, and $_{\leftarrow}$*ntnt* respectively. However, these distinctions are meaningful only in metatheories where Transitivity need not hold. Under a classical metatheory some definitions are coextensive.

9 Whether the fact that they have exactly the same collections of valid and invalid arguments is enough to say that they are the same logic is something we will not discuss here.

10 A formula A is a *logical truth* (in a logic **L** presented under a semantics S) if and only if, for each interpretation i (of S), $1 \in i(A)$.

11 As one reviewer pointed out, the resulting logic does not share the same meta-arguments only if we assume local validity. If we assume global validity, it also shares the same meta-arguments of classical logic and **ST**. For more information, see Barrio, Rosenblatt, and Tajer (2015).

of certain arguments, such as Transitivity, disjunctive and non-disjunctive logical consequence relations may be coextensive, as we will see later. However, one of the lessons we learn from tackling this problem is that presentation matters. In other words, the way we define logical consequence relations is important, so that our results do not appear *ad hoc*. Finally, we also show here that $_{\leftarrow}$*ntf* is useful so that *tonk* does not lead to triviality and can be considered as a meaningful connective.

Before concluding, a crucial point to address is the apparent equivalence between disjunctive and non-disjunctive definitions of logical consequence. A reviewer has suggested there would be no substantial differences between, for example, Cook's definition (using disjunctive clauses) and our non-disjunctive proposal, since both could be equivalent in a classical framework. However, this alleged equivalence depends on some arguments that are invalid in the logical consequence relations we analyze, such as *st* or $_{\leftarrow}$*ntf*. To clarify this, let us first consider the structure of the definitions.

Recall **TL**'s disjunctive definition of logical consequence. A is a logical consequence of Γ in **TL**, $\Gamma \vDash_{TL} A$, if and only if either: for every interpretation i, if $1 \in i(B)$ for all $B \in \Gamma$, then $1 \in i(A)$; or for every interpretation i, if $0 \in i(A)$, then $0 \in i(B)$ for some $B \in \Gamma$. Let A be "X is true"; B "Y is true"; C "Y is false"; and D "X is false". TL's logical consequence relation could be $(A \to B) \vee (C \to D)$. Now, consider the following equivalence proof between the two notions:

I. $(A \to B) \vee (C \to D)$	(TL's logical consequence)
II. $({\sim}A \vee B) \vee ({\sim}C \vee D)$	(I., Extensionality)
III. $({\sim}A \vee D) \vee (B \vee {\sim}C)$	(II., Commutativity and Associativity)
IV. ${\sim}({\sim}A \vee \mathrm{D}) \to (B \vee {\sim}C)$	(III., Extensionality)
V. $(A \wedge {\sim}D) \to (B \vee {\sim}C)$	(IV., de Morgan and Double Negation)

Considering the meanings of A, B, C, and D, logical consequence can be read as

st: An argument $\Gamma \vDash_{L} \Delta$ is logically valid if and only if, for every interpretation, if the premises are only true, the conclusions are not only false.

The proof of equivalence requires the following assumptions:

— $A \to B \dashv_{L} \vdash {\sim}A \vee B$	Extensionality
— $A \vee B \vdash_{L} B \vee C$	Commutativity
— $(A \vee B) \vee C \vdash_{L} A \vee (B \vee C)$	Associativity
— ${\sim}({\sim}A \vee B) \vdash_{L} (A \wedge {\sim}B)$	de Morgan
— ${\sim}{\sim}A \vdash_{L} A$	Double Negation
— If $A \vdash_{L} B$ and $B \vdash_{L} C$ then $A \vdash_{L} C$	Transitivity

While these definitions are equivalent in a classical metalanguage, such equivalence presupposes the validity of arguments, like Transitivity, that are invalid in the logical consequence relations we have outlined. Thus, the equivalence between disjunctive and non-disjunctive forms, which critically depends on this property. Finally, although the proof of the previous theorem in the metalanguage uses Transitivity, this does not guarantee that the defined relations (such as $_{\leftarrow}$*ntf*) inherit the same property. In fact, the non-transitivity of $_{\leftarrow}$*ntf* and *st* clearly exemplifies how we typically use a slightly stronger logic in the metatheory. In summary, the equivalence between disjunctive and non-disjunctive definitions appears to depend on Transitivity, at least in the standard proofs or frameworks considered here. However, it remains an open question whether such equivalence could be established independently of Transitivity.

4. Bonus

Mixed connectives such as *tonk* raise the question of whether they are of the same type as some of the connectives that were used to define them or whether they are of a new type. In the case of *tonk*, the question is whether it is a disjunction, a conjunction, both or neither, or perhaps it is a completely new type of connective. However, how do you know whether a connective is, say, a disjunction?

Before answering this question, we need to define some concepts. Following Estrada-González and Nicolás-Francisco (2024), in a Dunn semantics, an expression of the form $v_i \in (A)$ with $v_i \in \{1, 0\}$ is called a *Dunn atom*. Let $v_i \in (A)$ be a Dunn atom: we will say that $v_j \notin i(A)$, with $v_i, v_j \in \{1, 0\}$ and $v_i \neq v_j$, is its *Boolean counterpart*. For example, the following cases (considered horizontally) are Boolean counterparts of each other:

$0 \in i({\sim}A)$	$1 \notin i({\sim}A)$
$1 \in i(A \vee B)$	$0 \notin i(A \vee B)$

A *tweaking* is a modification in the evaluation conditions of a connective in which the only changes consist of substituting Dunn atoms for their Boolean counterparts.

Let us now consider the logic **FDE**. **FDE** can be presented by means of a language L, constructed in the usual way, with the following connectives: ~, ∧, ∨. The evaluation conditions and tables are the same as those presented in Section 2. The logical consequence relation of **FDE** is as follows: Let A and Γ be a formula and a set of formulas of L, respectively. A is a logical consequence of Γ in **FDE**, $\Gamma \vDash_{\mathbf{FDE}} A$, if and only if, for every interpretation i, if $1 \in i(B)$, for all $B \in \Gamma$, $1 \in i(A)$.

Estrada-González and Nicolás-Francisco (2024) say that a connective ⊚ is a *classically clear case* of negation/conjunction/disjunction/conditional, if:

1. The evaluation conditions of ⊚ are the negation/conjunction/disjunction/conditional conditions of **FDE**; or
2. The evaluation conditions of ⊚ are obtained from a *tweaking* of the evaluation conditions of **FDE**.

A connective ⊛ is a *negation/conjunction/disjunction/conditional* if and only if there is a clear case ⊚ of negation/conjunction/disjunction/disjunction/conditional such that:

$$\circledcirc(A_1,\ldots,A_n) \dashv_{\mathrm{L}}\vdash \circledast(A_1,\ldots,A_n)$$

We believe that few would doubt that the connectives ∨ and ∧ are classically clear cases of disjunction and conjunction, respectively. Although the tables with four interpretations look different from the classical ones, the evaluation conditions of ∨ and ∧ are the same. We can state that the disjunction ∨ and the conjunction ∧ satisfy 1. However, the evaluation conditions of ≷ do not correspond to those of any **FDE** connective and cannot be obtained by tweaking its evaluation conditions. Then ∨ and ∧ are classically clear connectives, unlike ≷.

Now let us consider a version of the **FDE** disjunction with the three admissible interpretations {1}, {1,0}, {0}[12], i.e.,

[12] This table corresponds to the disjunction of logics such as **LP**, when presented with Dunn semantics. Thus, it is easy to see that both classical logic, **FDE** and **LP** have the same evaluation conditions for their connectives, but differ in their tabular presentation.

$A \vee B$	{1}	{1, 0}	{0}
{1}	{1}	{1}	{1}
{1,0}	{1}	{1, 0}	{1, 0}
{0}	{1}	{1, 0}	{0}

It is easy to see that the following arguments are ${}_{\leftarrow}ntf$ valid

$$A \gtrless B \vdash_{\leftarrow ntf} A \vee B \text{ and } A \vee B \vdash_{\leftarrow ntf} A \gtrless B$$

That is, it holds that

$$A \gtrless B \dashv_{\leftarrow ntf} \vdash A \vee B$$

We can say that, in this approach, $\gtrless$ is a disjunction. Since the truth conditions are the same for $\vee$ and for $\gtrless$, the proof is trivial. Both $A \gtrless B$ and $A \vee B$ have only one interpretation where they are not true, $0 \in i(A)$ and $0 \in i(B)$. Hence, when $1 \notin i(A \vee B)$ then $0 \in i(A \gtrless B)$, and vice versa.

Now let us consider a version of the conjunction of **FDE** with the three admissible interpretations {1}, {1,0}, {0}[13] , i.e.,

$A \wedge B$	{1}	{1, 0}	{0}
{1}	{1}	{1, 0}	{0}
{1,0}	{1, 0}	{1, 0}	{0}
{0}	{0}	{0}	{0}

It is easy to check that the following arguments are ${}_{\leftarrow}ntf$-valid

$$A \gtrless B \vdash_{\leftarrow ntf} A \wedge B \text{ and } A \wedge B \vdash_{\leftarrow ntf} A \gtrless B$$

That is, it holds that

$$A \gtrless B \dashv_{\leftarrow ntf} \vdash A \wedge B$$

We can say that, in this approach, $\gtrless$ is also a conjunction. Since the falsity conditions are the same for $\wedge$ and for $\gtrless$, it is enough to look at the truth tables to see that $1 \notin i(A \wedge B)$ if and only if $1 \notin i(A)$ or $1 \notin i(B)$. Finally, in these same interpretations, $0 \in i(A \gtrless B)$. Then, $A \gtrless B \vdash_{\leftarrow ntf} A \wedge B$. On the other hand, $1 \notin \mathrm{i}(\mathrm{A} \gtrless B)$ only in the case where $0 \in i(A)$ and $0 \in i(B)$. In the same interpretation, $0 \in i(A \wedge B)$. Then $A \wedge B \vdash_{\leftarrow ntf} A \gtrless B$.

It is not strange that $\gtrless$ is a conjunction and a disjunction at the same time, given its introduction and elimination rules. However, in Estrada-González and Nicolás-Francisco's method, logical consequence plays a decisive role in the clarifi-

[13] This table corresponds to the conjunction of logics such as **LP**, when presented with Dunn semantics.

cation of this type of mixed connectives. For example, if we were to evaluate $A \lessgtr B \dashv_{\leftarrow ntf} \vdash A \vee B$, and $A \lessgtr B \dashv_{\leftarrow ntf} \vdash A \wedge B$ with a truth-preserving logical consequence relation, the result would be that $\lessgtr$ is only a disjunction, but not a conjunction. However, to show our result, we use homogeneously the same consequence relation of the theory, i.e. ${}_{\leftarrow}ntf$-validity. Therefore, $\lessgtr$ here is both a disjunction and a conjunction. The argument that *tonk* is a disjunction and a conjunction is not conclusive, certainly. This is because the conclusion depends on ${}_{\leftarrow}ntf$-validity, and although we have argued that this is a good notion to work with in this context, different logical consequence relations may give different results about what kind of connective *tonk* is.

5. Conclusion

In recent times, many people emphasize the limits of logic and formalization, appealing to the most diverse limitative theorems. With this paper we wanted to give an example that, in logic, with enough care and freeing oneself from some unnecessary assumptions, one can have the cake and eat it.

When many have considered a formal result to be definitive, they do so on the basis of various assumptions, which are often implicit. In the case of *tonk*, for example, the validity of Transitivity in the consequence relation was an implicit assumption that distorted *tonk*'s picture. Probably the lesson to be drawn here is that there are no strange or bad connectives per se, only connectives that may be incompatible with certain languages and certain consequence relations. This lesson is a generalization of Belnap's work which, as we said, suggests that *tonk* is a trivializing connective only in transitive logics.[14]

We showed that, without artificial motivation, it is possible to obtain a logical consequence relation under which *tonk* does not trivialize. Also, we proved that it is possible to define *tonk* with fewer admissible interpretations than those used by Cook. Finally, as a bonus, we presented a way of identifying *tonk* as a connective that is both a disjunction and a conjunction.

Acknowledgements

We would like to thank the members of Lógica M∃*x*∀ for their comments on previous versions of the typescript. The second author thanks especially Manuel Eduardo Tapia-Navarro and Dubian Cañas-Mora for their suggestions to improve some passages. We also thank the reviewers for their comments, which significantly improved the final version of the text. This work was supported by the projects CBF2023-2024-55 and PAPIIT IN406225.

REFERENCES

Barceló, A. A., (2008). Patrones inferenciales. *Crítica: Revista Hispanoamericana de Filosofía*, *40*(120), 3-35. https://doi.org/10.22201/iifs.18704905e.2008.993

Barrio, E., Rosenblatt, L., & Tajer, D. (2015). The logics of strict-tolerant logic. *Journal of Philosophical Logic*, *44*(5), 551-571. https://doi.org/10.1007/s10992-014-9342-6

[14] In the same spirit, Ramírez-Cámara and Estrada-González (2021) also argue that a "nasty" connective such as *knot* (which, unlike *tonk*, invalidates many classically valid argument schemes even though it looks like a mere generalized identity connective) is not so nasty in non-reflexive logics, either.

Belnap, N. D. (1962). Tonk, Plonk and Plink. *Analysis*, *22*(6), 130-134. https://doi.org/10.2307/3326862

Béziau, J.-Y. (2001). Sequents and bivaluations. *Logique et Analyse*, *44* (176), 373-394.

Buacar, N. M. (2018). Tonk y la concepción semanticista del significado. *Cuadernos de filosofía*, *70*, 5-18. https://doi.org/10.34096/cf.n70.6498

Carnap, R. (1943). *Formalization of logic*. Harvard University Press.

Chemla, E., Égré, P., & Spector, B. (2017). Characterizing logical consequence in many-valued logic. *Journal of Logic and Computation*, *27*(7), 2193-2226. https://doi.org/10.1093/logcom/exx001

Church, A. (1944). Review of *Formalization of Logic*, by R. Carnap. *The Philosophical Review*, *53*(5), 493-498. https://doi.org/10.2307/2181359

Church, A. (1953). Non-normal truth-tables for the propositional calculus. *Boletín de la Sociedad Matemática Mexicana*, *10*(1-2), 41-52.

Cobreros, P., Égré, P., Ripley, E., & van Rooij, R. (2012a). Tolerant, classical, strict. *Journal of Philosophical Logic*, *41*, 347-385. https://doi.org/10.1007/s10992-010-9165-z

Cobreros, P., Égré, P., Ripley, E., & van Rooij, R. (2012b). Tolerance and Mixed Consequence in the S'valuationist setting. *Studia logica*, *100*, 855-877. https://doi.org/10.1007/s11225-012-9422-y

Cook, R. T. (2005). What's wrong with Tonk (?). *Journal of Philosophical Logic*, *34*(2), 217-226. https://doi.org/10.1007/s10992-004-7805-x

da Ré, B., Szmuc, D., Chemla, E., & Égré, P. (2023) On Three-Valued presentations of classical logic. *Review of Symbolic Logic*, *16*(3) 1-23. https://doi.org/10.1017/S1755020323000114

Dunn, J. M. (1976). Intuitive semantics for First-Degree Entailment and "coupled trees". *Philosophical Studies, 29*, 149-168. https://doi.org/10.1007/BF00373152

Dunn, J. M. (2000). Partiality and its dual. *Studia Logica*, *66*, 5-40. https://doi.org/10.1023/A:1026740726955

Estrada-González, L. (2019). The Bochum Plan and the foundations of contra-classical logics, *CLE e-Prints*, *19*, 1-21.

Estrada-González, L. (2022). Dunn semantics for contra-classical logics, *EPTCS*, *358*, 298-309.

Estrada-González, L., & Nicolás-Francisco, R. A. (2024). Connexive negation. *Studia Logica*, *112*, 511-539. https://doi.org/10.1007/s11225-023-10078-x

Omori, H., & Sano, K. (2015) Generalizing functional completeness in Belnap-Dunn logic, *Studia Logica, 103*(5), 883-917. https://doi.org/10.1007/s11225-014-9597-5

Prior, A. N. (1960). The runabout inference-ticket. *Analysis*, *21*(2), 38-39. https://doi.org/10.2307/3326699

Ramírez-Cámara, E., & Estrada-González, L. (2021). Knot is not that nasty (but it is hardier than tonk). *Synthese*, *198* (22), 5533-5554. https://doiorg.pbidi.unam.mx:2443/10.1007/s11229-019-02498-x

Ripley, E. (2015). Anything goes. *Topoi*, *34* (1), 25-36. https://doi.org/10.1007/s11245-014-9261-8

Shoesmith, D., & Smiley, T. (1978). *Multiple Conclusion Logic*. Cambridge University Press.

Stevenson, J. T. (1961). Roundabout the runabout inference-ticket. *Analysis*, *21*(6), 124-128. https://doi.org/10.1093/analys/21.6.124

Teijeiro, P. (2020). Not a Knot. *Thought: A Journal of Philosophy*, *9*, 14-24. https://doi.org/10.1002/tht3.437

Wansing, H., & Shramko, Y. (2008). Suszko's thesis, inferential many-valuedness, and the notion of a logical system. *Studia Logica*, *88*(3), 405-429. https://doi.org/10.1007/s11225-008-9111-z

Wintein, S. (2016). On all Strong Kleene generalizations of classical logic. *Studia Logica*, *104*, 503-545. https://doi.org/10.1007/s11225-015-9649-5

Luis Estrada-González is a Mexican philosopher. He is a Permanent Researcher at the Institute for Philosophical Research of the National Autonomous University of Mexico (UNAM), and, since November 2024, Director of the Institute. Member of the Mexican National System of Researchers (SNII, level III), the Mexican Academy of Sciences, and the Logicians' Liberation League. Associate Editor of *Studia Logica* and member of the International Editorial Board of *Stoa*.

Address: Institute for Philosophical Research, National Autonomous University of Mexico, Circuito Mario de la Cueva s/n, Ciudad Universitaria, Mexico City, 04510, Mexico. – Email: loisayaxsegrob@comunidad.unam.mx – ORCID: https://orcid.org/0000-0002-1466-0240

Christian Romero-Rodríguez is a Colombian philosopher. Currently an Assistant Professor at the Universidad del Valle, in Cali, Colombia. He was a postdoctoral researcher in a SECIHTI project on logical pluralism, based at the Institute for Philosophical Research, UNAM. His research focuses on non-classical logics, especially paraconsistent and empty logics.

Address: Universidad del Valle, Ciudad Universitaria Meléndez, Calle 13 #100-00, Cali, Valle del Cauca, 760032, Colombia. – Email: christian.romero@correounivalle.edu.co – ORCID: https://orcid.org/0000-0002-7211-3392

Theoria, 2026, 41(1), 22-35
https://doi.org/10.1387/theoria.27206

'Ο Θεὸς ἀριθμετίζει
THEORIA
ISSN 0495-4548 – eISSN 2171-679X

On asking

(Sobre las preguntas)

Leonardo Flamini*
University of Johannesburg
https://orcid.org/0000-0001-8793-1921

Keywords

Genuine questions
Rhetorical questions
Inquiry
Information-seeking acts
Aim-constitutivist theory of questions
Inquiry-based sincerity conditions
Inquiring states of mind

ABSTRACT: Based on a survey testing people's intuitions about questions, Lani Watson has recently claimed that the act of seeking information captures the nature of genuine questions and what distinguishes them from rhetorical ones. In this paper, I will argue that Watson's account fails to provide an accurate theory of questions. By revisiting the survey cases and results, and using Searle's conception of the illocutionary point of directives, I will first argue that the cases support and confirm a simple and straightforward aim-constitutivist theory of the speech act of asking: One asks a real question by uttering an interrogative sentence only if one's utterance is an attempt that aims to get the addressee to answer the question it expresses. Otherwise, one is asking a fake or rhetorical question. Moreover, based on Friedman's perspective on inquiring attitudes as states in which one has a question open in one's thought, I will argue that the survey cases can be taken to expand and deepen the sincerity conditions for questions that Searle had initially proposed: One sincerely asks a question if one wants the addressee to answer the uttered question to settle an inquiring state of mind, or to make a target audience enter into it. Finally, in the cases in which the survey participants recognise the presence of a question despite the absence of any apparent speech act of asking, I will argue that they recognise an inquiring state of mind rather than simply an information-seeking act: They recognise that one is in a mental state in which one is asking oneself a particular question.

Palabras clave

Preguntas genuinas
Preguntas retóricas
Investigación
Actos de búsqueda de información
Teorías de las preguntas en términos de fines constitutivos
Condiciones de sinceridad basadas en la investigación
Estados mentales investigadores

RESUMEN: Basándose en una encuesta que comprueba intuiciones sobre preguntas, Lani Watson ha afirmado recientemente que el acto de buscar información captura la naturaleza de las preguntas genuinas, y lo que las distingue de las preguntas retóricas. En este artículo, argumento que la propuesta de Watson no consigue ofrecer una teoría de las preguntas adecuada. Mediante el examen de los casos y resultados de la encuesta, y recurriendo a la concepción de Searle del fin ilocutivo de los actos directivos, argumentaré primero que los casos respaldan y confirman una teoría simple del acto del habla de preguntar en términos de un fin constitutivo: solo se hace una pregunta genuina profiriendo una oración interrogativa si la proferencia es un intento orientado al fin de conseguir que el receptor conteste la pregunta expresada. De otro modo, estaremos haciendo una pregunta falsa o retórica. Además, basándome en la perspectiva de Friedman sobre las actitudes investigadoras como estados en los que hay una pregunta abierta en nuestro pensamiento, argumentaré que los casos de la encuesta pueden verse como expandiendo y profundizando las condiciones de sinceridad para preguntas propuestas inicialmente por Searle: hacemos sinceramente una pregunta si queremos que el receptor responda la pregunta proferida para cerrar una estado mental investigador, o hacer que la audience entre en un estado de este tipo. Finalmente, en los casos en que los participantes en la encuesta reconocen la presencia de una pregunta pese a la aparente ausencia de un acto del habla interrogativo, argumentaré que reconocen un estado mental investigador, más que un simple acto de búsqueda de información: reconocen que en uno está en un estado mental en el que se hace a uno mismo una pregunta particular.

* ***Correspondence to:*** Leonardo Flamini. African Centre for Epistemology and Philosophy of Science (ACEPS), University of Johannesburg, UJ on Empire, 5th Floor, Corner Barry Hertzog and Napier Road, 2092, Johannesburg, South Africa – leonardo.flamini3@gmail.com – https://orcid.org/0000-0001-8793-1921

How to cite: Flamini, Leonardo (2026). «On asking»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 41(1), 22-35. (https://doi.org/10.1387/theoria.27206).

Received: 11 January, 2025; Final version: 22 October, 2025.
ISSN 0495-4548 - eISSN 2171-679X / © 2026 UPV/EHU Press

Gako-hitzak

Benetako galderak
Galdera erretorikoak
Jakin-nahia
Informazioa bilatzeko egintzak
Galderen teoria xede-konstitutibista
Jakin-nahian oinarritutako zintzotasun-baldintzak
Jakin-nahizko egoera mentalak

LABURPENA: Pertsonek galderen inguruan dituzten intuizioak ebaluatzen dituen inkesta batean oinarrituta, Lani Watson-ek berriki adierazi du ezen informazioa bilatzeko egintza berak jasotzen duela zein den benetako galderen izaera eta zerk bereizten dituen galdera erretorikoetatik. Artikulu honetan, arrazoituko dut Watson-en proposamenak huts egiten duela galderen teoria egoki bat eskaintzeko saioan. Inkestako kasuak eta emaitzak berriro aztertuz eta Searle-k egintza direktiboen helburu ilokutiboaz duen ikusmoldea baliatuz, lehenik eta behin, argudiatuko dut kasu horiek galdetzearen hizketa-egintzaren teoria xede-konstitutibista sinple eta erraz bat babesten eta berresten dutela: galderazko perpaus bat egiten dugunean benetako galdera bat egiten ari gara, baldin eta gure enuntziatuaren helburua bada hartzaileak bertan planteatzen zaion galderari erantzutea; bestela, galdera faltsu edo erretoriko bat egiten ariko gara. Gainera, Friedman-ek jakin-nahizko jarrerei buruz duen ikuspegitik abiatuta (galdera ireki bat, hau da, erantzunik izan ez duen galdera bat darabilgu buruan egoera horietan), arrazoituko dut ezen inkestako kasuak baliagarriak direla Searl-ek galderetarako hasieran proposatu zituen zintzotasun-baldintzak zabaltzeko eta sakontzeko: galdera bat zintzotasunez egiten dugu, baldin eta gure nahia bada hartzaileak galdera horri erantzutea gure jakin-nahizko egoera mentala asebetetzeko edo entzulea egoera mental horretara eramateko. Azkenik, itxuraz galdetzearen hizketa-egintzarik egon ez arren inkestako parte-hartzaileak galdera baten presentziaz jabetu diren kasuetan, argudiatuko dut ezen, informazioa bilatzeko egintza soil bat baino gehiago, jakin-nahizko egoera mental bat hauteman dutela: nork bere buruari galdera jakin bat egiten ari zaion egoera mentalean daudela atzematen dute.

1. Introduction

What is a question? Lani Watson (2021) has recently answered this question by claiming that questions are *information-seeking acts*: When one asks a question, one seeks information, and "if something is an information-seeking act, then it gets a place in the collection" (2021, p. 293) of what we consider a question.[1] Moreover, based on this, she implies that this information-seeking act is what distinguishes utterances that are *genuine/real questions* from those that are *rhetorical/fake questions*. Specifically, from the perspective she proposes, it follows that one asks a real question by one's utterance when and only when one is seeking information. Instead, when asking a question rhetorically or pretending to ask a question by one's utterance, one is not seeking information.[2]

Watson arrives at this intuitive conclusion by analysing the results of a survey that tests people's intuitions about questions in everyday scenarios.[3] One of the central survey results is that while most participants recognise the presence of questions in the first two cases, they do not detect any real questions in the last two scenarios:

> ROAD
>
> Sarah is trying out a new route to school. Along the route she comes to the side of a busy, unfamiliar road with no pedestrian crossing. She looks both up and down the road before crossing to check if there are any vehicles coming and then proceeds to cross safely. (Watson, 2021, p. 281)

---

1 Watson acknowledges that there already are logical and linguistic theories of questions (for example, see Ciardelli *et al.*, 2019; Fiengo, 2007; Groenendijk & Stokhof, 1997; Krifka, 2011; and Wiśniewski, 2013). However, she argues that these logical and linguistic accounts involve such technical and formal machinery that makes them distant from what we usually consider a question. Given this, Watson proposes shifting our attention from what a question is from a linguistic or logical point of view to what we do in practice when we raise questions.

2 In what follows, I will adopt the following terminology: I use the terms "fake questions" and "rhetorical questions" interchangeably to refer to utterances that, although they take the linguistic form of a question, do not constitute instances of the speech act of asking. Conversely, I use the expressions "genuine questions" and "real questions" to refer to utterances that, in their constitution, do amount to the speech act of asking (see Section 3 for further discussion).

3 You can fill out the survey at www.philosophyofquestions.com/questionnaire/.

> COUNTRIES
>
> Sarah, is listening to a colleague describing a lesson he has just given on countries of the world. While he is talking, Sarah realises that she doesn't know how many countries there are and, as she is interested to know, she interjects saying 'how many countries are there'. Various colleagues respond with several different figures. (Watson, 2021, p. 285)
>
> DISINTERESTED
>
> Sarah has arrived at school and is walking to her classroom down a long corridor. She spots a particularly disliked colleague approaching from the other end. As they pass, despite having no interest whatsoever in his wellbeing, Sarah glances up and mutters 'morning, how are you'. (Watson, 2021, p. 279)
>
> RAIN
>
> Sarah has just woken up and opened the curtains to discover that it is raining again, as it has been for the past two weeks. On seeing this she exclaims out loud 'will it ever stop raining'. Her partner, who is still in the bed behind her, shrugs. (Watson, 2021, p. 279)[4]

Based on this result, Watson first concludes that questions do not need to be *interrogative sentences* or *speech acts*. Namely, questions do not need to be *linguistic entities*. Indeed, she notes how the survey participants recognise the presence of a question in scenarios like ROAD, where no question is linguistically articulated.[5] Very intuitively, Watson explains this result by claiming that the survey participants identify the presence of a question in COUNTRIES and ROAD, although there is an interrogative speech act in COUNTRIES while there is none in ROAD, because they recognise the presence of an *information-seeking act*. Indeed, Sarah seeks whether any vehicles are coming when she looks up and down the road in ROAD. Moreover, she seeks how many countries there are when she asks her spoken question in COUNTRIES. Instead, Watson's explanation for why survey participants do not identify real questions in DISINTERESTED and RAIN is that they recognise that Sarah is not seeking information when she utters her interrogative sentences. Indeed, Sarah's question in DISINTERESTED seems just a way to respect a formal kind of politeness required in the workplace: Sarah, being totally disinterested in her colleague's well-being, is not really seeking to know anything about it. Moreover, Sarah's question in RAIN appears to be just a way to vent her frustration towards the weather: She is not really seeking to determine whether the rain will ever stop since it is evident that, sooner or later, it will.

Based on this information-seeking-act-based reading of the survey results and on the assumption that the survey participants correctly recognise when there is a genuine or a rhetorical question in the cases they are confronted with, Watson concludes that the act of seeking information captures the nature of genuine questions and what distinguishes them from rhetorical ones. Namely, it captures when a question is genuinely or rhetorically asked.

In this paper, despite the previous intuitive conclusion, I will argue that Watson's account fails to capture the nature of genuine questions and what distinguishes them from rhetorical ones. In Section 2, we will see how Watson solves a pos-

---

[4] In response to ROAD, "66% of the survey participants judge there to be a question (3,670 participants), 28% judge there to be no question (1,589 participants), and 6% are unsure (352 participants)" (Watson, 2021, p. 281). In response to COUNTRIES, "95% of the survey participants judge there to be a question (4,711 participants), 3% judge there to be no question (147 participants), and 2% are unsure (94 participants)" (Watson, 2021, p. 285). In response to DISINTERESTED, "38% of the survey participants judge there to be a question (1,955 participants), 54% judge there to be no question (2,804 participants), and 8% are unsure (407 participants)" (Watson, 2021, p. 279). In response to RAIN, "35% of the survey participants judge there to be a question (2,115 participants), 58% judge there to be no question (3,432 participants), and 7% are unsure (406 participants)" (Watson, 2021, p. 279).

[5] As the reader can note, in all cases Watson proposes to the survey participants, there are no *question marks*. However, the survey participants have no problems identifying questions in the previous scenarios. Watson argues that this is a "further point in favour of a definition of questions that does not rely on their manifestation in language" (2021, p. 283).

sible counterexample to her theory. I will point out that, even if her solution seems to protect her account, it actually generates other counterexamples to her own theory of questions, rendering the resulting theory unable to distinguish genuine from rhetorical questions.

However, in Section 3, by revisiting the survey cases and results and adopting Searle's (1969, 1979) and Searle and Vanderveken's (1985) idea of the *illocutionary point of directives*, I will show how Watson's survey might be used to support and confirm a simple and straightforward *aim-constitutivist theory of the speech act of asking*: One asks a real question by uttering an interrogative sentence only if one's utterance is an attempt that aims to get the addressee to answer the question it expresses.[6] Instead, a question is rhetorically asked when one is not making the previous *aim-directed attempt* by one's interrogative sentence.

Moreover, based on Friedman's (2013, 2017, 2019) perspective on inquiring attitudes as states in which a question is open in one's thought, I will argue that the survey cases can be taken to expand and deepen the *sincerity conditions* for questions that Searle (1969) had initially proposed: One sincerely asks a question if one wants the addressee to answer the uttered question to settle an inquiring state of mind, or to make a target audience enter into it. Said informally, using the most common inquiring attitudes human beings can have, a question is sincere when it is an act of will to solve or generate *curiosity* or a *wondering state*.

Finally, in the cases in which the survey participants recognise the presence of a question despite the absence of any speech act of asking or any interrogative sentence being uttered, I will argue that they recognise an inquiring state of mind rather than simply an information-seeking act: They recognise that one is in a mental state in which one is asking oneself a particular question.

In a nutshell, I will argue that the survey results can be seen as indicating that we recognise the existence of questions as speech acts, with their constitutive and sincerity conditions, and as mental states.

2. *What goes wrong with Watson's account?*

Consider the following case:

> EXPOSE
>
> Sarah is attending a staff meeting in which a colleague she doesn't particularly like (the one that provoked the disinterested "how are you" from earlier) is being unhelpfully rude and obstructive regarding a particular issue. Sarah knows that he has not read the minutes from last week's meeting, which he did not attend, and in order to expose this she interjects at the end of his comments, saying, "what did you think of the suggestion Julia made last week to address this issue". (Watson, 2021, p. 289)

For this scenario, the survey participants agree that there is a question.[7] However, it appears that Sarah is not seeking information by asking her question. Indeed, she already knows well that her colleague has not read the minutes from last week's meeting. In light of this, one might argue that EXPOSE represents a counterexample to Watson's theory: Not all questions are information-seeking acts.

---

[6] Note the distinction between questions as *speech acts* and as the *semantic content* that is expressed by one's speech act or interrogative utterance. A question as a speech act refers to a type of act done by our language, whereas a question as semantic content refers to the abstract informational structure typically expressed by an interrogative (see footnote 1 for accounts of questions as semantic entities).

[7] In response to EXPOSE, "62% of the survey participants judge there to be a question (3,038 participants), 29% judge there to be no question (1,391 participants), and 9% are unsure (430 participants)" (Watson, 2021, p. 289).

Watson recognises the problem. However, she holds that Sarah is asking a genuine question, and she is indeed seeking information in EXPOSE. Specifically, Watson claims that Sarah is exploiting the *function of questions* of seeking information to achieve her *practical purposes*: Exposing and embarrassing her disliked colleague to make him stop being unhelpfully rude and obstructive. Namely, from Watson's perspective, Sarah decided to seek information from her colleague by using her question to achieve her goal. However, she notes that Sarah could follow another strategy. For example, to obtain the same result, she could make an assertion like: "Stop being so rude and obstructive regarding this particular issue: You didn't even read the minutes from last week's meeting!".

Given this reading of EXPOSE, Watson can defend her position against a possible counterexample that individuates a genuine question that is not an information-seeking act. Indeed, against the potential objector, Watson claims that the information-seeking act that defines questions does not need to be determined by an *information-seeking motive*: It is not always directed at satisfying one's informational needs. Rather, it is an act that can be used for multiple purposes, such as the *practical one* of exposing and embarrassing a disliked colleague to make him stop being unhelpfully rude and obstructive. Moreover, by having this response available, Watson can still maintain her explanation as a correct account of what occurs when survey participants identify a question: They recognise an information-seeking act.

2.1. Watson's solution generates counterexamples to her own theory of questions

Watson's solution might appear intuitive and correct at first sight. However, it turns out to be a double-edged sword for her own theory of questions. Indeed, notice that the same argument Watson uses for reading EXPOSE as a case in which we have an information-seeking act can be extended to DISINTERESTED and RAIN: Scenarios in which the survey participants do not recognise any questions, and in which Watson claims there are rhetorical ones.

Indeed, by using Watson's own argument for EXPOSE, one might say that Sarah is seeking information in DISINTERESTED because, despite being totally disinterested in knowing about her colleague's well-being, she is using the information-seeking function of her question, "How are you?", to attain the goal of respecting a formal kind of politeness required in the workplace. Namely, in DISINTERESTED, Sarah would have decided to seek information from her colleague by her interrogative utterance to attain the previous practical goal. Moreover, one might also say that Sarah is seeking information in RAIN because, although she is not really aiming to know whether the rain will stop, she is using the information-seeking function of her question, "Will it ever stop raining?", just to vent her frustration towards the weather. Namely, in RAIN, Sarah would have decided to seek information from her partner by her interrogative utterance to express her frustration. In a nutshell, we can see that, by applying Watson's argumentative strategy from EXPOSE, we can say that Sarah's interrogative utterances in DISINTERESTED and RAIN are information-seeking acts.

However, it is easy to see how this reading of DISINTERESTED and RAIN, as a direct application of Watson's solution, would negate her own theory of questions. On the one hand, since her framework states that being an information-seeking act is a sufficient condition for something to be an instance of a question, then the interrogative utterances Sarah raises in DISINTERESTED and RAIN would be real cases of questions. However, as the survey participants and Watson herself recognise, there is no question in the previous scenarios: Sarah is not really asking a question with her interrogative utterances but rather posing a rhetorical or fake one. As a consequence, the application of Watson's argumentative strategy from EXPOSE to DISINTERESTED and RAIN would generate *counterexamples* to her own theory of questions: It would show that being an information-seeking act is not always sufficient for something to be a question. In particular, considering the previous scenarios, such an act would not be enough to make an interrogative utterance a real case of what we consider a question. Therefore, contrary to her own theory, Watson's solution would imply that not all information-seeking acts fall under the collection of what we consider a question.

On the other hand, the previous reading of DISINTERESTED and RAIN would also indicate that not only genuine but also rhetorical questions can be information-seeking acts. Indeed, in DISINTERESTED and RAIN, there would be questions we intuitively conceive as rhetorical or fake that are information-seeking acts. As a consequence, the application of Watson's strategy from EXPOSE to DISINTERESTED and RAIN would generate counterexamples to her own conception of rhetorical questions as acts that do not seek information. Moreover, given this, it would imply that both genuine and rhetorical questions can function as information-seeking acts. Indeed, applying Watson's solution, both the real questions in cases like COUNTRIES and the rhetorical questions in DISINTERESTED and RAIN would be information-seeking acts. However, if this is true, then a simple information-seeking act could no longer serve as the distinguishing criterion between genuine and rhetorical questions, being a feature that both of them can share.

Finally, given this last point, Watson's strategy from EXPOSE applied to DISINTERESTED and RAIN would show that an information-seeking act is present in both cases where the survey participants recognise questions and in scenarios where they do not. Given this common feature between the scenarios, it would follow that the presence or absence of an information-seeking act, as defined by Watson, cannot be the contrasting feature that the survey participants intuitively and correctly recognise when they distinguish the presence or absence of a question. Rather, this commonality of the cases would suggest that the survey participants detect something beyond the simple act of seeking information when they discriminate whether Sarah is asking a question or not. Therefore, the outcomes generated by applying Watson's strategy from EXPOSE would indicate that her account cannot accurately explain what is going on with the survey participants' intuitions. As a consequence, the same participants' intuitions and responses could not serve as reliable support for her information-seeking-act-based theory of questions: The survey results, in themselves, would not support the idea that a question is an information-seeking act.

In conclusion, the solution Watson provides to explain why there is an information-seeking act in EXPOSE would prima facie protect her account. However, upon closer examination, we can see that it would generate counterexamples to her theory of questions, render her theory incapable of distinguishing genuine from rhetorical questions, and ultimately make her account fail to explain the survey results —thereby making them unsuitable as support for her own theory.

3. *Watson's survey reconsidered: Questions as speech acts and mental states*

The problems in Watson's account push us to find an alternative answer to the questions: What does Watson's survey reveal regarding participants' intuitions about questions? What does it show about a theory of questions, particularly regarding the distinction between genuine and rhetorical questions?

In this last section, I reconsider first the survey cases in which a question is verbally uttered, drawing on Searle's (1969, 1979) and Searle and Vanderveken's (1985) idea of the *illocutionary point of directives*. Based on this, I will argue that the participants' responses support a simple *aim-constitutivist theory of the speech act of asking*: One asks a real question by uttering an interrogative sentence only if one's utterance is an attempt that aims to get the addressee to answer the question it expresses. In other words, for an interrogative utterance to count as a real question, it must have *the goal* of getting the addressee to provide a statement about the relevant question expressed by the uttered interrogative sentence. Or, put differently, it must aim to make the addressee describe how things are regarding the pertinent question. Given this perspective, we can also distinguish rhetorical or fake questions from genuine ones: If one's interrogative utterance does not amount to this *aim-directed act*, then it is a rhetorical or fake question.

Moreover, considering Friedman's (2013, 2017, 2019) conception of *inquiring attitudes* as mental states in which one has a question open in one's thought, I will further argue that the survey cases can be taken to show some *inquiry-based sincerity conditions* for the speech acts of asking. These conditions expand and deepen Searle's original theorisation: One asks a sincere question by one's interrogative utterance if one wants the addressee to answer the uttered question to resolve an inquiring attitude or to generate it in a target audience. Said more intuitively by adopting the most common inquiring attitudes we can have, a question is sincere when, by one's interrogative utterance, one wants to solve one's curiosity or wondering state, or to generate it in a target audience. In a nutshell, from this perspective, a question is sincere if it is the act of will to solve an inquiring state of mind or to generate it in a target audience. Conversely, if a question is not sincere, one is not doing this act of will.

Finally, considering this, I will argue that in survey cases where no question is verbally articulated or performed as a proper speech act, the survey participants recognise the presence of questions because they recognise that one is in an inquiring attitude. In a nutshell, the survey can be taken to show that the *act of asking* can be a *speech act* with some constitutive and sincerity conditions or simply a *mental state* by which we open a question in our thoughts.

3.1. Questions as speech acts: Aim-constitutivism

In Searle's (1969, 1979) famous perspective on speech acts, the *illocutionary point* (or what he also calls the *essential condition*) is what helps us individuate and define the different types of illocutionary acts we can perform with our language. In Searle and Vanderveken (1985), we find the following definition: "The illocutionary point of a type of illocutionary act is that purpose which is essential to its being an act of that type" (1985, p. 14). In other words, it is the aim that is *constitutive* of the illocutionary act: If an utterance X amounts to the illocutionary act of the type T, then X aims at I. Conversely, if one's utterance X does not aim at I, then X is not an illocutionary act of the type T. In a nutshell, the illocutionary point is the *constitutive aim* of the different types of illocutionary acts that tells us something about the very nature of each of them.[8]

In Searle's (1979) and Searle and Vanderveken's (1985) well-known taxonomy, questions are a type of *directive*. Directives are those illocutionary acts that are defined by the following illocutionary point: They are attempts by the speaker that aim at getting the addressee to do something. Hence, if an utterance amounts to a directive, then that utterance has the aim of getting the addressee to do something. In the specific case of questions, the speaker's utterance aims to get the addressee to answer the question it poses. Hence, following the aim-constitutivist perspective derived from Searle's original idea of the illocutionary point, an utterance counts as the illocutionary act of asking only if it aims to get the addressee to answer the question it expresses. If the utterance does not have this aim, it does not constitute the illocutionary act of asking. In other words, one is not really or genuinely asking a question if one's utterance does not amount to the previous *aim-directed act*. Notably for our purposes, this aim-constitutivist theory of the speech act of asking provides a basis for distinguishing between real/genuine and fake/rhetorical questions: If one's interrogative utterance is a genuine or real question, then it has the aim constitutive of the illocutionary act of asking. If it lacks this aim, then the interrogative utterance is a rhetorical or fake question.

Now, it is interesting to note that Watson's cases and survey responses appear to support and confirm this account and distinction when a given question is verbally articulated to an addressee in the various scenarios. Indeed, consider the following cases in addition to COUNTRIES and EXPOSE:

> GOOGLE
>
> Sarah, is attempting to find out where the nearest butchers to her house is, on behalf of a friend. As she is a vegetarian, Sarah herself has no interest in this information but, nevertheless, she types 'local Edinburgh butchers' into Google and notes down the location. (Watson, 2021, p. 278)
>
> PENCILS
>
> Sarah is teaching her students basic arithmetic. One student shows her his answer to a question and she sees that it is wrong. In order to correct him, rather than telling him the answer, Sarah counts out ten pencils on the table in front of him, removes two and says 'how many pencils are left on the table'. (Watson, 2021, pp. 287-288)

8 See Alston (1991), Searle (1991), and Siebel (2002) for critical discussions on the concept of illocutionary point.

In all these cases, the survey participants affirm that there is a question, and it is indeed intuitive to think that Sarah is asking real questions with her sentences in these scenarios.[9] Relevantly, comparing these cases and in line with our theory, we can appreciate that what they have in common is this minimum relevant fact: By her utterance, Sarah is trying to get the addressee to answer the question it expresses. Specifically, she attempts to elicit a statement from the addressee regarding the question she verbalises. Or, put differently, she aims to make her addressee provide a description of how things are regarding the question expressed in her utterance.

Indeed, in COUNTRIES, through her interrogative sentence, Sarah is trying to get the expert to answer how many countries there are. In EXPOSE, she is trying to make her disliked colleague provide a statement regarding her question, namely that he did not read the minutes from last week's meeting. In GOOGLE, even if there is no human addressee but a digital one like Google, Sarah's written utterance can be seen as a shortcut interrogative sentence for the complete one, "Where are the local butchers in Edinburgh?": By posing this interrogative to Google, she tries to make it provide some results that describe where to find the local butchers. Finally, in PENCILS, by her interrogative sentence, Sarah is trying to make her student answer how many pencils are left on the table. In other words, all these cases confirm and support what is explained by our aim-constitutivist theory of the speech act of asking: Sarah's utterances are questions, as our intuitions suggest, and they are indeed aimed at getting the addressee to answer the question each of them expresses. Hence, in accordance with the survey participants' responses, our theory provides an explanation of why these cases involve the presence of questions: Sarah is simply performing the *speech act of asking* by her interrogative utterances.

Instead, in cases where the survey participants do not recognise a real question, it is notable that Sarah's utterance is not aimed at getting the addressee to answer the question by making a statement or describing how things are. Indeed, in DISINTERESTED, Sarah's interrogative sentence is not really aimed at getting her colleague to answer and describe how he is doing: Being totally disinterested in it, her question is just a formal courtesy that does not require any answer in exchange. The same happens in RAIN. The interrogative sentence Sarah utters is not aimed at getting her partner to answer the question expressed and make a statement about whether the rain will ever stop. Sarah's question is just a way to express her frustration about the weather. Given our aim-constitutivist theory, Sarah is not asking questions in DISINTERESTED and RAIN through her interrogative utterances because they lack the illocutionary point of the illocutionary act of asking. Rather, since these interrogative utterances are not aimed at getting the addressee to answer them, they represent cases of fake or rhetorical questions. Hence, in line with the survey participants' responses, our theory effectively explains why these cases do not involve the presence of questions, even though interrogative sentences are spoken aloud. Simply, Sarah is not performing the *speech act of asking* by her interrogative utterances.

Therefore, considering our analysis of the previous cases, we can see that the illocutionary point of the utterance performed in the scenarios proposed by Watson enables us to distinguish between cases of real and fake questions. In particular, we can appreciate that *when* survey participants recognise the presence of a question in scenarios where an interrogative sentence is verbally articulated, the interrogative utterance is aimed at getting the addressee to answer the question it expresses. Instead, the participants do not recognise the presence of a real question *when* the interrogative utterance does not have this aim. Given this, it follows that in cases where a question is verbally uttered, the survey participants' responses track this distinctive feature rather than simply the information-seeking act Watson describes. In other words, it is simply reasonable to conclude that, in these cases, participants recognise the presence or absence of the speech act of asking, as defined by the theory of this linguistic act we have proposed. Ultimately, given this, the survey cases and responses can be taken to confirm and support the aim-constitutivist theory of questions we have retrieved from Searle, along with the distinction it yields between genuine and rhetorical questions.

---

[9] In response to GOOGLE, "72% of the survey participants (to date) judge there to be a question (3,515 participants), 20% judge there to be no question (996 participants), and 8% are unsure (365 participants)" (Watson, 2021, p. 278). Finally, in response to PENCILS, "84% of the survey participants judge there to be a question (4,223 participants), 12% judge there to be no question (583 participants), and 4% are unsure (215 participants)" (Watson, 2021, p. 288).

3.2. Questions as speech acts: Sincerity conditions

Beyond these conclusions, I believe the survey scenarios not only support the previously established aim-constitutivist conception of the speech act of asking but also clarify its sincerity conditions. Namely, they reveal not just what constitutes a genuine question when verbally uttered, but also what makes it sincere.

In Searle's perspective, when one performs an illocutionary act, one expresses a *psychological state*. To be sincere in an illocutionary act, the speaker must have the same psychological state expressed by the illocutionary act the speaker performs. In the case of directives, they express a psychological state that can be associated with *wants* or *desires*: The speaker wants the addressee to do what she is attempting to get the latter to do (Searle, 1969, 1979; Searle & Vanderveken, 1985). Therefore, to have a sincere directive, the speaker must have the corresponding want or desire. Given this perspective, which characteristic want or desire is expressed by the speech act of asking? Following this account and considering that asking is a type of directive, the simple answer is that asking is sincere only if *one wants* or *desires the addressee to answer the question one expresses by one's utterance*. However, I think that, taking into account the survey cases, we can be more detailed about the types of wants that are characteristic of the speech act of asking. Indeed, intuitively, in all the cases we have considered in the previous subsections, Sarah wants the addressee to answer the question expressed by her utterance. But, to understand better the type of want involved in the speech act of asking, we should ask: Why does Sarah want the addressee to answer?

Considering this question, it is worth noting that one type of want that Searle recognises for having a sincere question is that one wants the addressee to answer the question because the speaker wants to find out and, therefore, know the answer.[10] Now, going beyond Searle's terminology and drawing on contemporary epistemology literature, this want that is expressed by the speech act of asking can be more precisely understood or explained by a type of attitude Friedman (2013, 2017, 2019) characterises as an *inquiring state of mind*: A state of mind in which one has a question open in one's thought and aims to answer it. Generalising from this perspective, we have the following insight: One asks a sincere question through one's interrogative utterance when one wants the addressee to answer the question it expresses because one is in an inquiring state of mind towards it.[11] Or, put more concisely, one asks a sincere question through one's interrogative utterance when one is in an inquiring state of mind towards the question expressed by the utterance.

Indeed, following this view, in COUNTRIES, it appears that Sarah wants the addressee to answer because she is in a *psychological state* in which she is asking herself the relevant question and aims to know the answer: She has the question, "How many countries are there?", open in her thought and aims to find out the answer. In particular, in COUNTRIES, Sarah seemingly exhibits a specific variety of the inquiring attitudes Friedman generally describes: She has a *state of curiosity*. As a matter of fact, when posing her interrogative sentence, Sarah can be simply described as being curious about how many countries there are and to know the answer. Therefore, following our intuitions, her question appears sincere because she is in a state of curiosity about it. In other words, using Friedman's terminology, she wants the addressee to answer the question, "How many countries are there?", because she has an inquiring attitude towards it.

Turning to GOOGLE, it is clear that Sarah is neither personally curious about the question of where the local Edinburgh butchers are nor genuinely interested in knowing the answer for herself. However, despite this lack of personal curiosity, she can still be understood as embodying an inquiring attitude when she poses her question to Google. Indeed, Sarah does aim to find out the answer and resolve the question of where the local Edinburgh butchers are. Given the details of the case, we can imagine that she is in this condition because her not-so-sensible non-vegetarian friend asked her a favour, as they do not know anything about Edinburgh. In this intuitive description of the case, although she is not personally invested, Sarah is nonetheless wondering about the previous question when she submits her written interrogative to Google. Therefore, even without personal curiosity, her question can be considered sincere because

[10] See Whitcomb (2017) and Haziza (2023) on the norms governing interrogative speech acts that express inquiries.

[11] See Drucker (2022) for a similar account that is focused on the state of wondering.

it reflects a *form of wondering* —a type of inquiring attitude— that, in the case of Sarah, is pragmatically or instrumentally motivated given her situation: She keeps the question open in her mind and seeks to settle it, not for herself but to assist her friend. Again, given this perspective, Sarah's question can be taken as sincere because it expresses the psychological state of a person in an inquiring state of mind, and Sarah herself does occupy a form of inquiring attitude.

Now, considering PENCILS and EXPOSE, it is obvious that Sarah is not in an inquiring state of mind about the questions she asks her student or disliked colleague: She already knows the answers and is totally settled about them. However, despite this, I believe that the questions uttered can be regarded as sincere under specific conditions that are always related to an inquiring attitude. Indeed, in both cases, I think Sarah wants to make a target audience —whether her student or colleagues— enter into an inquiring state of mind regarding the question expressed by her interrogative utterance, with the ultimate goal of making them know the answer.

Starting with PENCILS, by her interrogative, Sarah can be seen as attempting to make her student answer the question of how many pencils are left for the following reason: She wants him to consider that question reflectively and try to answer it by himself, thereby learning the result of ten minus two and exercising his mathematical skills. Intuitively, given this description of the case, even if Sarah is not inquiring into the question she poses, she still appears to ask her student a genuine and sincere question —provided she asks it with the kind of want we have just outlined. From this perspective, the question is sincere because, even if Sarah does not have an inquiring attitude towards it, she wants her student to enter into this very state. Put differently and more intuitively, using the common inquiring attitude of curiosity or wonder, we can describe her speech act of asking as sincere for the following reason: By her question, she wants her student to become curious or start wondering about it so that he can try to answer it by himself and come to know the answer.

Before discussing EXPOSE, it is relevant to note that there is another notable way, acknowledged by Searle (1969), by which Sarah's *teaching* or *exam question* might be sincere: She might want the student to answer her question because she wants to know whether he knows the answer.[12] Even if we can interpret Sarah's question in this way in PENCILS, it is important to underline that an inquiring attitude always appears to be involved and can be taken to explain Sarah's want. Indeed, Sarah can be described as having an inquiring attitude towards whether her student knows the answer to her question of how many pencils are left on the table: She is wondering about this, being a teacher who must evaluate and teach her students. In this eventuality, Sarah asks her student the question of how many pencils are left on the table to solve the inquiring attitude she has as an evaluating teacher. In other words, she wants her student to answer her question because, as Searle notes, she wants to know whether the student knows the answer. However, more specifically, this want amounts to or is explained by Sarah's state of wondering about whether her student knows the answer. In conclusion, this is a further way to expand Searle's theorisation, in the case of teaching and exam questions, by explaining the want that makes a speech act of asking sincere through an appeal to an inquiring attitude: One's question is sincere because one does want the addressee to answer the question due to one's inquiring state of mind towards whether the addressee knows the answer.

Finally, moving to EXPOSE, this case can be intuitively accounted for in a way similar to our first description of PENCILS. Indeed, Sarah can be taken to sincerely ask the question because she wants her colleagues to enter into an inquiring attitude towards the same question she poses: She wants them to consider it and become curious about the answer. In particular, she does this to make them learn that the disliked colleague did not read last week's meeting minutes. Indeed, in support of this, it is intuitive to imagine that if the disliked colleague refused to answer Sarah's question or stayed silent about it, the curiosity of the other colleagues would increase so that they themselves would require an answer from him. In this eventuality, due to the curiosity instilled by Sarah's initial question in her colleagues, the disliked colleague would then be forced to confess that he did not read the minutes of last week's meeting. Therefore, from this perspective, we can intuitively see Sarah's question as sincere, even if she is not inquiring about it due to her knowledge, because she wants her col-

12 See also Gaszczyk (2025) on the speech act of exam questions.

leagues to enter into an inquiring attitude towards it and make them know the truth. In a nutshell, following Searle's necessary condition for a sincere question, Sarah's question is intuitively sincere in PENCILS and EXPOSE because she wants the addressee to answer the question. However, more specifically, she wants this and thus asks a sincere question as she wants the target audience to enter into an inquiring state of mind regarding the same question she poses.

In conclusion, by analysing the survey cases, we can see that we can theoretically expand the sincerity conditions for the speech act of asking initially described by Searle. Generally, one asks a sincere question by one's utterance only if one wants the addressee to answer the relevant question. However, more specifically, one's question is sincere if one wants the addressee to answer the question to solve an inquiring attitude, or to make a target audience enter into it. In a nutshell, a question is sincere if it is the act of will to solve an inquiring state of mind or to generate it in a target audience. If a question is not sincere, then the speaker is simply not performing this act of will.

3.3. Questions as mental states

Reading the previous two subsections, one might note that, even though our account explains why the survey participants recognise the presence of questions when a question is verbally articulated in an utterance and a speech act is performed, it remains silent on another important point. Specifically, it does not explain why participants also recognise the presence of questions in scenarios where a proper speech act of asking does not take place. Indeed, the account proposed is a theory of the speech act of asking, which, intuitively, does not cover cases where there are no speech acts.

However, despite this possible concern, I think that the account we offered can be used to explain why the survey participants recognise a question in cases where there does not seem to be a proper speech act of asking. In order to show this, consider ROAD and the following scenarios:

> DICTIONARY
>
> Sarah, has arrived in her classroom and is checking through her teaching resources for the day. She notices a word that she is not familiar with and, in order to find out what it means, she looks it up in a dictionary. (Watson, 2021, p. 281)
>
> SKY
>
> Sarah is leaving work for the day and night has already fallen. She pauses at one point and looks up to the clear, dark sky above. Marvelling at the scale and beauty of the scene she says silently inside her head 'how big is the universe'. She considers this for several minutes and then continues home. (Watson, 2021, p. 293)

The survey participants say there are questions in all of these cases.[13] The question is now: Why?

Watson argues that Sarah is performing an information-seeking act in all the previous scenarios. Hence, based on this, her explanation was that the survey participants recognise a question in the previous cases because they recognise this information-seeking act. Even if this description seems to fit what is happening in the cases, we can retrieve a much simpler and more fundamental description from our theory of the speech act of asking. Indeed, one might ask the following question: "It is true that Sarah is seeking information in all these cases. However, again, what explains why she is seeking information?". Intuitively, the simple answer is that Sarah is *asking herself* a question: "Are any vehicles coming?", "What is the meaning of this unfamiliar word?", and "How big is the universe?". Namely, adopting Friedman's

---

13 In response to DICTIONARY, "81% of the survey participants judge there to be a question (4,119 participants), 14% judge there to be no question (725 participants), and 5% are unsure (260 participants)" (Watson, 2021, p. 281). In response to SKY, "53% of the survey participants judge there to be a question (2,592 participants), 35% judge there to be no question (1,698 participants), and 12% are unsure (560 participants)" (Watson, 2021, p. 293).

terminology, the previous questions are open and unsettled in Sarah's mind: She has an inquiring attitude towards "Are any vehicles coming?", "What is the meaning of this unfamiliar word?", and "How big is the universe?".

Therefore, even if it might be correct to describe Sarah as doing an information-seeking act in the previous cases, it is a more robust and intuitive description that Sarah is simply in a mental state in which she is asking herself a given question —she is in an inquiring state of mind. Given this perspective, we can draw the following conclusion about the survey participants' intuitions: What they recognise in ROAD, DICTIONARY, and SKY when they say that there is a question is not simply the action of seeking information, but more profoundly that Sarah is in an inquiring state of mind in which she has the relevant question open in her thought. Put more informally, the survey participants recognise that Sarah is asking herself a question due to her curiosity or state of wondering about it. In a nutshell, we can say that, rather than simply recognising an information-seeking act, the survey participants recognise the presence of a *mental state of asking,* which motivates and guides Sarah's information-seeking behaviour.

Before concluding, I would like to add something more about the case SKY. Indeed, our description aligns well with the case if we assume that Sarah, with her silent question in her head, is not performing any *proper speech act*. However, I recognise that one could object that Sarah's question is *internally verbalised* and that she is doing an *inner speech act* by silently posing her question in her head. That said, even if we want to grant this, our theory can still account for this case and why there is a question. Indeed, we can intuitively say that Sarah is performing an *inner speech act of asking*. In particular, following the account drawn from Searle, Sarah is performing an internal variant of the linguistic act of asking, in which the addressee from whom she attempts to get the answer through her internal interrogative utterance is herself. In other words, this case can be taken to show that the illocutionary point of the illocutionary act of asking is preserved even when the speaker and the addressee coincide: By making her *inner utterance*, Sarah is attempting to get herself to answer the question expressed by her innerly uttered interrogative sentence.

Moreover, we can also account for the intuitive sincerity of her question. Always following our theory, we can say that Sarah's inner question is also sincere because she wants herself to answer it due to her inquiring attitude: She has the question, "How big is the universe?", open in her thought. Indeed, as we noted, Sarah intuitively wonders and is curious about how big the universe is. Nonetheless, we can go even beyond this simple and intuitive explanation. Based on our theory, we could also intuitively say that Sarah's inner question is sincere because, by silently uttering it in her mind, she wants to put herself in a *reflective inquiring state of mind* in which she thinks about the previous question and explores it with awareness. Namely, by making her inner utterance and aiming to get herself to answer the question, she wants to wonder in a reflective way about the question of how big the universe is.

In conclusion, putting together all our considerations about Watson's survey, we can conclude that it shows the following: We recognise the existence of questions not only as *speech acts* but also as *mental states*. In the first case, by following Searle's theorising about speech acts, we have argued that the survey cases confirm and support an aim-constitutivist theory of the speech act of asking: One's utterance is a genuine question only if it is aimed at getting the addressee to answer the question it expresses. Namely, one's utterance constitutes a real question only if it amounts to this aim-directed act. Moreover, upon further analysis of the survey scenarios, we have expanded and deepened Searle's original sincerity conditions for the speech act of asking: One sincerely asks a question if one wants the addressee to answer the uttered question to settle an inquiring attitude, or to make the target audience enter into it. Finally, in the second case, by following Friedman's conception of the inquiring state of mind, we have argued that the survey cases suggest that there is a type of asking that is not linguistic but mental: One can be in a state of mind in which one has a question open in one's thought. Put differently, one can be in a state of mind in which one is asking oneself a question.

Acknowledgements

I thank the two anonymous reviewers for their helpful comments and suggestions, which contributed to improving this work.

REFERENCES

Alston, W. P. (1991). Searle on illocutionary acts. In E. Lepore & R. van Gulick (Eds.), *John Searle and his critics* (pp. 57-80). Blackwell.

Ciardelli, I., Groenendijk, J., & Roelofsen, F. (2019). *Inquisitive semantics*. Oxford University Press. https://doi.org/10.1093/oso/9780198814788.001.0001

Drucker, D. (2022). Wondering on and with purpose. In U. Kriegel (Ed.), *Oxford studies in philosophy of mind* (Vol. 2, pp. 58-84). Oxford University Press. https://doi.org/10.1093/oso/9780192856685.003.0003

Fiengo, R. (2007). *Asking questions: Using meaningful structures to imply ignorance*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199208418.001.0001

Friedman, J. (2013). Question-directed attitudes. *Philosophical Perspectives*, 27(1), 145-174. https://doi.org/10.1111/phpe.12026

Friedman, J. (2017). Why suspend judging? *Noûs,* 51(2), 302-326. https://doi.org/10.1111/nous.12137

Friedman, J. (2019). Inquiry and belief. *Noûs*, 53(2), 296-315. https://doi.org/10.1111/nous.12222

Gaszczyk, G. (2025). Interrogatives, inquiries, and exam questions. *Erkenntnis*, 90(4), 1437-1456. https://doi.org/10.1007/s10670-023-00758-4

Groenendijk, J., & Stokhof, M. (1997). Questions. In J. van Benthem & A. ter Meulen (Eds.), *Handbook of logic and language* (pp. 1055-1124). Elsevier. https://doi.org/10.1016/B978-044481714-3/50024-2

Haziza, E. (2023). Questioning and addressee knowledge. *Synthese,* 201(4), 1-23. https://doi.org/10.1007/s11229-023-04113-6

Krifka, M. (2011). Questions. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 2, pp. 1742–1785). De Gruyter Mouton. https://doi.org/10.1515/9783110255072.1742

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139173438

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press. https://doi.org/10.1017/CBO9780511609213

Searle, J. R. (1991). Response: Meaning, intentionality, and speech acts. In E. Lepore & R. van Gulick (Eds.), *John Searle and his critics* (pp. 81-102). Blackwell.

Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. Cambridge University Press.

Siebel, M. (2002). What is an illocutionary point? In G. Grewendorf & G. Meggle (Eds.), *Speech acts, mind, and social reality* (Vol. 79, pp. 125-139). Springer. https://doi.org/10.1007/978-94-010-0589-0_9

Watson, L. (2021). What is a question. *Royal Institute of Philosophy Supplement*, 89, 273-297. https://doi.org/10.1017/S1358246121000114

Whitcomb, D. (2017). One kind of asking. *The Philosophical Quarterly,* 67(266), 148-168. https://doi.org/10.1093/pq/pqw027

Wiśniewski, A. (2013). *Questions, inferences, and scenarios*. College Publications.

Leonardo Flamini is a Research Associate at the African Centre for Epistemology and Philosophy of Science (ACEPS), University of Johannesburg. He specialises in epistemology, with a focus on the theory of inquiry.

Address: African Centre for Epistemology and Philosophy of Science (ACEPS), University of Johannesburg, UJ on Empire, 5th Floor, Corner Barry Hertzog and Napier Road, 2092, Johannesburg, South Africa. E-mail: leonardo.flamini3@gmail.com – ORCID: https://orcid.org/0000-0001-8793-1921

Theoria, 2026, 41(1), 36-56
https://doi.org/10.1387/theoria.27048

THEORIA

ISSN 0495-4548 – eISSN 2171-679X

Perceptual reference, object files and Molyneux's Question

(Referencia perceptual, archivos de objetos y la cuestión de Molyneux)

Sergio Cermeño-Aínsa*
University of Girona
https://orcid.org/0000-0002-9598-8386

Keywords

Molyneux Question
Perceptual Shape Representation
Cross-modal Perception
Object Files

ABSTRACT: Molyneux's Question (MQ) —whether a newly sighted might immediately recognize tactilely familiar shapes by sight alone— remains inconclusive. The most common way to pose the question is in representational terms, i.e., whether vision and touch generate or not similar representational types or whether there is some intrinsic similarity between visual and tactual shape representations. Recent developments in cross-modal perception suggest that if visual and tactile representations are linked, then a positive response to MQ is the most probable conclusion. In this paper, I explore and provide some suggestions for this possibility. In particular, I give rise to the object file strategy, according to which, if part of the tactile information is referentially encoded and amodally stored in object files, then the category-specific contents of R(vision) and R(touch) are in some way propositionally (conceptually) linked, thus opening a wide avenue to think that a newly sighted might recognize tactilely familiar shapes by sight alone.

Palabras clave

Cuestión de Molyneux
Representación Perceptual de las Formas
Percepción Intermodal
Archivos de Objetos

RESUMEN: La Cuestión de Molyneux (MQ por sus siglas en inglés) —si un recién vidente podría reconocer visualmente formas previamente conocidas solo mediante el tacto— sigue sin ser concluyente. La forma más común de plantear esta cuestión es en términos representacionales, es decir, si la visión y el tacto generan tipos de representación similares o si existe alguna similitud intrínseca entre las representaciones visuales y táctiles en el reconocimiento de las formas. Avances recientes en la percepción intermodal sugieren que, si las representaciones visuales y táctiles están vinculadas, una respuesta positiva a MQ es la más probable. En este artículo, exploro y presento algunas ideas que sugieren esta posibilidad. En particular, planteo la estrategia de los archivos de objetos, según la cual, si parte de la información táctil se codifica referencialmente y se almacena amodalmente en archivos de objetos, entonces los contenidos específicos de la categoría de R(visión) y R(tacto) estarán de alguna manera vinculados proposicionalmente (conceptualmente). Esto abre una amplia posibilidad para pensar que una persona recién vidente podría reconocer visualmente formas previamente conocidas solo mediante el tacto.

* ***Correspondence to:*** Sergio Cermeño-Aínsa. Department of Philosophy, Facultat de Lletres, University of Girona, Edifici Sant Domènec I, Pl. Ferrater i Mora, 1, Campus Barri Vell, Girona 17004, Spain – elsergei@hotmail.com – https://orcid.org/0000-0002-9598-8386

How to cite: Cermeño-Aínsa, Sergio (2026). «Perceptual reference, object files and Molyneux's Question»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 41(1), 36-56. (https://doi.org/10.1387/theoria.27048).

Received: 03 November, 2024; Final version: 16 July, 2025.

ISSN 0495-4548 - eISSN 2171-679X / © 2026 UPV/EHU Press

Gako-hitzak

Molyneux-en galdera
Formen irudikapen pertzepziozkoa
Pertzepzio crossmodala
Objektu-fitxategiak

LABURPENA: Molyneux-en galderak (MQ, ingelesezko siglen arabera) —ea ikusmena berreskuratu berri duen pertsona batek ikusmenaren bidez soilik berehala ezagut ditzakeen aldez aurretik ukimenaren bidez ezagun zitzaizkion formak— ez du oraindik behin betiko erantzunik. Ohikoena da galdera hori irudikapenaren ikuspegitik planteatzea; hau da, ea ikusmenak eta ukimenak antzeko irudikapen-motak sortzen dituzten, edo ea ba ote dagoen berezko antzekotasunik formen ikusmenezko eta ukimenezko irudikapenen artean. Pertzepzio crossmodalaren arloan egin berri diren aurrerapenei jarraikiz, ikusmenezko eta ukimenezko irudikapenak lotuta badaude, arrazoizkoena litzateke MQ galderaren erantzuna baiezkoa izatea. Artikulu honetan, aukera hori iradokitzen duten ideia batzuk aztertzen eta aurkezten ditut. Zehazki, objektu-artxiboen estrategia planteatzen dut. Estrategia horri jarraikiz, ukimenaren bidez eskuratutako informazioaren zati bat modu erreferentzialean kodetzen bada eta «objektu-artxiboetan» modu amodalean biltegiratzen bada, orduan R (ikusmena) eta R (ukimena) kategorien eduki espezifikoak lotuta egongo dira proposizionalki (kontzeptualki), eta horrek bide ematen du pentsatzeko ezen ikusmena berreskuratu berri duen pertsona bat gai izango dela ikusmenaren bidez soilik ezagutzeko ukimenaren bidez lehendik ezagun zitzaizkion formak.

1. Introduction

Molyneux's Question (MQ)[1] asks whether a person who has recently gained the ability to see would be able to recognize familiar tactile shapes, such as spheres or cubes, by sight alone. There are three possible responses: negative, positive and plural. Negative answers usually appeal to perceptual learning; simply, learning to see a cube is fundamentally different from learning to feel a cube. This is because the connections between our ideas of different shapes (cube and sphere) acquired through different sense modalities must be learned through direct experience; i.e., sensory ideas of sight and touch are inherently different (Locke, 1694; Berkeley, 1709; Condillac, 1754; Lotze, 1887). Arguably, negative responses have dominated the scene, and it is perhaps the first impression for everyone who has been in touch with MQ. Furthermore, negative responses have been supported, though not firmly, by empirical research.[2] As such, the burden of proof falls on those who argue for an affirmative answer.

Therefore, the challenge lies more in how to stand up for a compelling positive answer than in offering more arguments for the negative one. Affirmative answers have usually appealed to some kind of nativism concerning the mechanisms for conceptual acquisition. Different arguments have tried to substantiate the affirmative option. The first and most general suggests that subjects imaginatively construct a visual representation from their previous tactile experience (Leibniz, 1760). Current arguments, however, are more specific and amply supported by progress in developmental science and neuroscience. Noë (2004), for example, thinks that sensorimotor contingencies in sight and touch can share a common spatial content. Campbell (1996) appeals to the geometrical properties of objects, which constitute subjects' sensory experience, and makes the perception of shape by sight and touch similar in structure. Evans (1985) argues that our perception of the sphere and the cube are the same for touch and sight because both share the same egocentric relations, i.e., the ability to egocentrically localize the parts of shapes is alien to the sense modality in which it is perceived. For Held (2009), there must be a transference from tactile to visual representations via supra-modal pathways. And similarly, Pascual-Leone and Hamilton (2001) talk about meta-modal brain centres that perform particular computational operations without specific reference to any type of sensory modality.[3]

---

1 For a recent recompilation of essays exploring the long-standing issues that MQ presents to philosophers and psychologists, see Ferreti and Gleney (2021).

2 The first experimental data concerning MQ dates back to Cheselden (1728), who published an account of what a congenitally blind person had seen after his cataracts were removed. More recently, there have been several attempts to approach the issue empirically, such as Gregory and Wallace (1963), Fine *et al.* (2003), and especially Held *et al.* (2011), who conducted what has been the most discussed experiment to date. As we will see, this evidence is not free of controversies.

3 For other positive arguments, see Prinz (2002) or Levin (2008).

Beyond the yes/no responses there is still a third possibility. It has also been proposed that a monolithic answer to MQ is too simplistic; the complexity of MQ involves more than one problem and, therefore, requires more than one solution. This is the pluralist view. Reid (1764), for example, argued that MQ is ambiguous since a positive or negative response depends on whether the shapes are two or three-dimensional as well as on the expertise of the subjects. Matthen and Cohen (2020) claim that MQ comprises a cluster of subproblems, each of which requires a particular solution. And Nanay (2020) thinks that a compelling answer is only possible by considering the specificities of the type of blindness in question. In particular, whether blindness involves or not the impairment of visual mental imagery.

In this paper, I argue that it is possible to defend an affirmative response by employing what I call the object file strategy (OFS). OFS is consistent with some of the positive answers reviewed above; it shares many things with what Held (2009) called transference from tactile to visual representations via supra-modal pathways, and it is markedly aligned with what Pascual-Leone and Hamilton (2001) refer to as meta-modal brain centres that perform particular computational operations without specific reference to any type of sensory modality. However, as we will see, OFS has its proper specificities.

But let's start from the beginning. Any affirmative response to MQ requires understanding seeing and touching in representational terms. The question is: Do vision and touch generate (or not) similar representational types? Put differently, is there an intrinsic similarity, a common code, between visual and tactual shape representations? Suppose F is a shape property represented by R. Three possibilities deploy here:

1. If R(vision) and R(touch) are *identical*, then a newly sighted might recognize tactilely familiar shapes by sight alone.
2. If R(vision) and R(touch) are *referentially linked*, then a newly sighted might recognize tactilely familiar shapes by sight alone.
3. If R(vision) and R(touch) are *structurally linked*, then a newly sighted might recognize tactilely familiar shapes by sight alone.[4]

These are some of the possible positive responses defended by theorists. Not many see it as viable defending (1); there is no easy way to support that R(vision) and R(touch) produce the same types of representations. I think this option depends on what we understand by identical representations; basically on whether these similarities are in perceptual shape representations or recognitional shape representations. The latter may be identical since the concept of 'sphere' is the same for any sense modality, but this is not necessarily the case for the former (for discussion, see Green, 2022). Many others defend different versions of (2) or (3). One version of (2) was initially advocated by Evans (1985). Particularly, Evans assumes that if "the tactual concept is the same as the visual concept", then the answer to MQ is 'yes' (p. 381). And versions of (3) are recently supported by Cheng (2019). Option (3) depends on the existence of a tactile field structurally similar to the visual field (Cheng, 2019; Serrahima, 2023). However, there is no easy way to explain that both fields (tactual and visual) are identical beyond their referential link. For example, just as tactile experiences of external objects are mediated by interoceptive bodily sensations, visual experiences are purely exteroceptive (Martin, 1993; O'Shaughnessy, 1989). So, I propose to go on with (2) here. Option (2) puts the accent on a referential link between the two representations, but it does not require that the representations are identical or structurally linked. Specifically, I suggest that by leaning on a particular characterization of the notion of object file, it is possible to take a step toward a plausible variant of (2). My core claim reads as follows:

> **Object Files Strategy**: If part of the tactile information is *referentially* encoded and *amodally* stored in object files, then the *category-specific* contents of R(vision) and R(touch) are in some way *propositionally (conceptually) linked*, and therefore a newly sighted might recognize tactilely familiar shapes by sight alone.

[4] Of course, if R(vision) and R(touch) are not identical and do not have any conceptual or structural link, then a newly sighted will not recognize tactilely familiar shapes by sight alone.

OFS admits a difference between tactile and visual representation —they are not necessarily similar nor structurally linked but only referentially bonded. This makes OFS more in line with option (2).[5] In what follows, I will defend this position.

The paper is structured as follows. Section 2 provides some clarifications, specifications and particularities of MQ. This preliminary section pretends to clarify the nature of MQ and its peculiarities. Section 3 critically reviews some classical and novel experimental attempts to solve MQ. Section 4 focuses on recent developments in cross-modal perception. I argue that any positive answer to MQ requires (though does not exhaust) the existence of cross-modal perception. Section 5 presents and defends Object Files as the mechanisms for harbour cross-modal perceptual recognition. Section 6 deploys the Object File Strategy by presenting the properties of object files: referentially (subsection 6.1), amodality (subsection 6.2), propositionally (subsection 6.3) and category-specifically (subsection 6.4), and defends that, thus seen, OFS is a plausible view. Section 7 offers empirical evidence favouring the OFS, first from the presence of Multimodal Object Files (subsection 7.1) and second from experimental research on artificial agents and non-human animals (subsection 7.2). Section 8 concludes and provides some suggestions for future approaches.

2. *Preliminaries: specificities of MQ*

Before delving into the core of the question, let´s take a closer look at the particularities of the thought experiment suggested by Molyneux more than three centuries ago.

First, MQ only applies to visuo-tactile interactions in shape perception. MQ presents a situation where a blind person observes, for the first time, stimuli only touched before. In this situation, the property involved is ´shape´. Shape properties are directly processed by sight, touch or both —no one can smell, taste or hear shapes. One can perhaps do it indirectly, e.g., when I hear a horn, I recognize a car, perhaps the shape of a car, but in no way this constitutes a direct perception. Other properties involve other senses: colour is exclusive to vision, temperature to touch, sweetness to taste, loudness to hearing and floralness to odour. Some perceptual properties are also perceptible across different senses, sometimes providing overlapping information for objects and scenes in the environment. Cross-modal perception appears very early in infants (Purpura *et al.*, 2018; Joanne *et al.*, 2014) and continues developing throughout our lives (Norman, 2006).

It would not make much sense to question whether a subject who suddenly recovers from congenital deafness would recognize his mother's voice, nor whether a subject who suddenly recovers from congenital blindness could visually recognize a rose previously smelled. The answer to these questions is clearly negative,[6] which shows that, at least intuitively, the case posed by MQ reveals a particular conjunction between vision and touch in recognizing objects from shape properties. The specificity of shape for vision and touch is, therefore, essential to understanding MQ, which means that any response to MQ will not be, in principle, generalizable to the relationship between other sensory modalities. Note also that other properties like size are also perceptible through vision and touch, but size is not the type of property that produces the typical controversies that the MQ original case of shape produces. It seems evident that a newly sighted can visually differentiate between a basketball and a golf ball previously touched. Similarly, the difference in weight between the sphere and the cube (imagine the sphere weighs ten times as much as the cube) can only be appreciated by touch but not by sight. Consequently, much caution when extending the MQ case to any possible combination between other senses or different properties than shape.[7] MQ is, therefore, a very specific case.

---

5 Although OFS says nothing about (1) and (3), it can still be adjusted to both by introducing minor modifications. After all, the only difference between (1), (2) and (3) is where we put the accent, whether on identity, referentiality or structure. Although (1) and (3) are, I think, harder to defend, I will not argue here against them; I simply assume that OFS is more consistent with (2).

6 Cases of synesthesia are rare and will not be considered here.

7 This is crucial for the visual-centric bias in perception research (for interesting insights into why vision dominates perception research, see Hutmacher, 2019). The idea that everything we know about vision extends to other sensory modalities falls apart.

https://doi.org/10.1387/theoria.27048

Second, MQ only works in the appropriate direction. The reverse situation, i.e., if a person who can visually distinguish between a cube and a sphere would recognize it solely by touch without any prior tactile experience, seems less controversial.[8] Hardly one can replicate a case like this, but the reverse option offers quite a different impression. In this case, the affirmative response is much more applicable than the negative one. It can be argued that when visually perceived, the differences between stimuli are captured and registered with greater precision and strength than when perceived tactilely. Most likely, this is because visual recognition produces a more extensive imprint than tactile one. The reverse situation has already been explored in the field of robotics. Falco *et al.* (2019) showed that robots are able to recognize through tactile exploration previous visual exposition to objects without having touched any object before. Again, this is an unsurprising discovery since visual information is, by far, more informative than tactile one, at least regarding object shape recognition. However, the positive response to the reverse situation opens the possibility of deeply exploring a positive response to the original characterization. All this shows that the specificity of MQ requires examining the visuotactile multimodality in the direction suggested by the original case, but also that the mechanism by which visual shape information is transferred to the tactile one in the reverse situation can also work for the original case. Perhaps, the prevalence of vision over touch clouds the informational transfer from touch to vision, but not from vision to touch. In subsequent sections, this will be further explored.

Third, MQ involves a single content delivered in different formats. Indeed, it is easy to see that the same content can be manifested in various formats. For example, the content "that is a cube" can be represented in spoken English, written Spanish, illustrated in sign language, codified in Morse or even photographically. In principle, the same occurs in the opposite direction —different contents can be depicted in the same format. For example, the same artwork may have different content for an expert than for a non-expert, or the same word can have different contents, e.g., homonymous words. So, content and format are both parts of a representation, but they can be individuated.

The question raised by Molyneux places us in the first situation. According to MQ, similarities in representation might imply similarities in representational content but not in representational format. That is, the same content can be depicted in different formats, visually (iconically) or tactile (sensitively), but just as the representational content of seeing a sphere and the representational content of touching a sphere are similar, the format in which such content is represented differs. Therefore, in the MQ case, visual and tactile representations are felt phenomenally different because they are conveyed in different representational formats, but coincide in their representational content.[9] So, any positive answer to MQ must focus their arguments on the similarities of representational content.

To sum up this section: 1) MQ is a specific case involved in visuotactile interactions not extended to other perceptual systems, 2) MQ only works in the direction offered by the original case (the blind person that suddenly sees) when we take the reverse direction (the person who previously couldn't feel touch and suddenly gains the ability to do so), it does not work, 3) any positive response to MQ should focus on the content rather than the format in which such content stores in memory.

Now, once paved the ground let me show how there can be a compelling interpretation of option 2.

3. Can we solve MQ empirically?

Arguably, MQ hardly yields a clean and non-controversial experimental situation. The main reason is that the sudden restoration of sight with no other concomitant issues may not be realistic. The effects of congenital blindness on

---

8 Surprisingly, researchers have not taken this option into account. I have only read this variant in Bruno and Mandelbaum (2010), who also show astonishment about the lack of attention to the reverse case (see footnote 27).

9 I am not denying that phenomenological similarity is important for the categorization and identification of objects themselves (there are strong empirical reasons to think like that) but in my view, the particularities of the MQ case require, for a positive answer, focusing on coincidences in representational content (for a more nuanced discussion see Di Stefano and Spence, 2024a).

the mechanisms of vision (optical or cognitive) require significant time for adaptation from sight restoration to object shape recognition.[10] This objection has been put in different ways by past philosophers (La Mettrie, 1750; Smith, 1795) and continues to be the most significant objection to empirically testing MQ (Degenaar, 1996; Meltzoff, 1993; Cattaneo and Vecchi, 2011).

Despite this, much of the recent interest in MQ is motivated by the possibility of testing it experimentally. Technical and experimental advances urge us to reconsider old questions with new tools, and MQ is one of them. There are relatively few studies of subjects recovering from congenital blindness, which undoubtedly is a symptom of the scarce experimental guarantees traditionally offered about MQ. However, studies of subjects recovering from cataracts stand as one of the most promising experimental situations to address MQ.

An often-cited and much-discussed study conducted by Held *et al.* (2011) was thought, at least for a while, to have resolved the question towards a negative answer. These researchers examined five patients soon after cataract surgery for congenital blindness. The task involved two steps. In the first one participants were visually or tactilely presented with one target stimulus (Lego blocks). The second step was to identify from two further stimuli, again tactilely or visually, which one is identical to the target stimulus. Subjects' performance was compared across three conditions: the visual-visual condition, which of the seen shapes is identical to the seen target one; the tactual-tactual condition, which of the felt shapes is identical to the felt target one; and the tactual-visual condition, which of the seen shapes is identical to the target felt. The results indicated that in the two first conditions (visual-visual and tactual-tactual), subjects performed well —suggesting they were able to see and feel the objects— but their performance in the visual-tactile condition was very poor, approximately at chance. The experimenters concluded that the newly sighted subjects did not exhibit an immediate transfer of their tactile shape knowledge to the visual domain (p. 552), and, therefore, the answer to the Molyneux question is likely negative. However, and leaving aside important methodological concerns (see Connolly, 2013; Schwenkler, 2012, 2013; Cheng, 2015; Clarke, 2016; Levin, 2018), the experiment does not solve the pressing concern of the sudden restoration. In fact, experimenters observed that subjects could not visually recognize the stimuli within 48 hours after surgery, but they recognized them a few days after recovering their sight. Accordingly, the answer to MQ is negative immediately after sight restoration but turns positive after a short period.

In sum, the experimental literature discussing these issues has not offered, to date, a convincing response to the issues surrounding MQ. For many, the possibility of complying with the necessary conditions to run an experiment that could provide a persuasive response to MQ is remote (Cheng, 2015). Given all this, MQ appears as a thought experiment that, *at least in adult human subjects*, hardly can be transformed into a testable one. The experiment requires an abrupt recovery of the visual system, which is not very realistic in practice (although empirically possible). So, despite efforts to move the thought experiment into the realm of experimental practice, pending the definite experiment (if workable), any answer to MQ will ultimately require a certain amount of inference.

Alternatively, empirical research concerning MQ has also been developed in non-humans. It is worth highlighting, therefore, two lines of research that could be especially promising. The first comes from robotics (Falco *et al.*, 2019; Liu *et al.*, 2019), and the second from animal experimentation (Solvi *et al.*, 2020; Versace *et al.*, 2024). All these lines result in an affirmative answer to MQ and will be deeply addressed later.

---

[10] Almost from its inception, philosophers have suggested modifications to the original formulation. Diderot (1749) suggested using two-dimensional shapes to avoid initial failures in the visual processing of tridimensionality, Reid (1764) recommended using geometricians as subjects, and Leibniz (1760) suggested offering the subject some clues about the type of forms that must identify. More recently, Levin (2008) suggests providing subjects enough time for complete visual restoration while cancelling their ability to identify shapes by sight alone. Be as it may, and in light of this concern, MQ remains as follows: If the Molyneux subject were able to make a sufficient 'visual discrimination' between shapes, would she be able to recognize which is the cube and which is the sphere?

4. Evidence on cross-modal perception

Much of the positive answers to MQ are based on recent developments in cross-modal object recognition. Cross-modal object recognition is defined in neuroscience and psychology as *the ability to recognize an object, previously inspected with one modality e.g., touch, via a second modality, e.g., vision, without prior training in the second modality.*[11] An initial and fundamental step towards a positive answer to MQ is understanding that the different sensory modalities usually interact, overlap, modulate, integrate or benefit each other to capture the most appropriate interpretation of perceptual representations. This is often called intermodal perception. Extensive evidence shows that interactions between the senses are robust, abundant, varied and occur at a very early stage of perceptual processing (O'Callaghan, 2015). Of course, there are perceptual properties whose representational content is exclusive of a particular sensory modality —e.g., there is no such thing as perceiving colours through other modalities than vision— but the interaction between two or more sensory modalities is a very common situation for the representation of the vast majority of object properties (such as shape properties, perceptible by both vision and touch). There are two types of inter-modal effects in perception: multimodal perception (or multimodal integration) and cross-modal perception. The former occurs when two or more sensory modalities fuse to detect or identify stimuli (e.g., audio-visual speech), and the latter when an object initially perceived by one sense modality is recognized by another sense modality. Clearly, MQ falls in the second type of effect.

The concurrence of vision and touch in shape perception is not controversial; however, a response to MQ goes beyond the cross-modality occurrence. Affirmative and negative answers can easily accommodate this evidence —MQ does not question whether there is informational transmission from touch to vision but whether the learned tactual information is visually available even without visual information ever entering the retina before. Generally, studies on cross-modal perception do not consider these effects in blind people since visuo-tactile cross-modality is available to normal sighted subjects but not to blind ones. That is, cross-modal recognition is amply demonstrated for subjects with intact vision and touch, but MQ requires that cross-modal recognition exists in the absence of previous cross-modal experience; the specificity of the MQ case forces us to reconsider this issue deeply.

One potential solution to this issue is to consider that perceptual information is translated into a common code (Altieri, 2015), i.e., that there is a shared space for both sensory modalities in which the information is commonly encoded. This seems to occur in crossmodal transfer between haptic and visual representations. Although haptic and visual channels are independent in the earlier levels of processing, at the later levels, their signals are commonly encoded (Held, 2009). The lateral occipital complex (LOC), a region of the ventral visual pathway, has been proposed as the brain area where visual-haptic signals concur. This typical visual area not only responds to haptic 3-D and 2-D stimuli (Amedi *et al.*, 2002; Prather *et al.*, 2004) but also processes shape information independently of the sensory modality used to acquire it (Lacey and Sathian, 2015).[12]

Now, what type of information is commonly encoded in crossmodal perception? Let´s consider two possibilities. The first is that the incoming bottom-up structural information of the stimuli transfers from one sense to another. The second is that previously stored top-down conceptual knowledge transfers from one sense to another. The first option is not very realistic since vision and touch use different sense channels to process the structural information of objects. Recall that vision and touch share informational units in the later levels but not in the earlier ones (Held, 2009). The second is simply unsustainable since in the absence of incoming information we cannot speak of perception properly. In

---

[11] "Psychology dictionary," http://psychologydictionary.org/.

[12] The most recent studies in neuroscience show interactions between vision and touch in the cerebral cortex (Sathian and Lacey, 2022). These crossmodal interactions occur, for example, when haptic stimuli activate visual brain regions (Snow *et al.*, 2014), thus converging visual and haptic stimuli. These convergences occur in the association brain areas, mostly in the Lateral Occipital Cortex (LOC). These typically vision-specific areas are activated in the absence of retinal input (Amedi *et al.*, 2001) and encode haptically perceived shapes even for congenitally blind subjects (Amedi *et al.*, 2010). Moreover, LOC plays a key role in both visual and tactual object recognition (James *et al.*, 2006), suggesting that it integrates the information processed in both modalities.

my view, cross-modal perception requires the conjunction of both types of transmission: crossmodal transference of some structural properties of objects and crossmodal transference of the previously stored knowledge. Let me illustrate it better. The tactile system uses a generic structural shape property (e.g., the typical roundness of spheres) to ground the perceptual identification (i.e., sphere), and fundamentally, this is achieved by way of perceptual reference. Referential information acts, therefore, as the conjunction of the bottom-up incoming information and the top-down stored one. This is what is transferred from one sense to another, regardless of which one of the two types of encodings (bottom-up or top-down) acquires more weight in stimuli identification, the central point is that perceptual reference encloses the two encodings into a single one. Therefore, if we understand referentiality as the causal relation between referents-in-the-world and mental representations that refer to them, then becomes reasonable considering that it is the confluence between both types of informational units (i.e., referentiality) what it is transferred from one sense to another. In sum, when two mental representations co-refer, they represent the same object or property (Rescorla, 2020).

Let´s put all this in MQ terms. Perceptual reference requires both the ability to perceptually distinguish between objects (cube and sphere) and the ability to recognize or identify each one of them. On the one hand, subjects can differentiate between objects just by responding merely to differences in the incoming low-level spatial features without involving identification —subjects can find differences between the two objects without knowing which is which. On the other hand, subjects cannot identify stimuli whose properties do not arrive at their senses —perceptual identification cannot float free from low-level structural features. Thus, only if the informational exchange between tactile and visual perceptual systems is referential the newly sighted subject will be able to recognize crossmodally tactilely familiar shapes by sight alone. In short, subjects learn to differentiate tactilely between the sphere and the cube by creating a referential correspondence between the salient properties of each (e.g., cube with squareness and sphere with roundness) with the correspondent identificative label attached, thus referentially anchoring the cube and the sphere with their salient properties. Now, this referential anchorage should be registered through learning in order to be subsequently recovered through memory and, therefore, recognized, i.e. there will not be recognition without previous referential anchorage.

Suppose this story is correct. Suppose the MQ subject visually distinguishes between the sphere and the cube because her brain referentially matches the top-down conceptual information previously registered by touch with the bottom-up spatial information incoming by sight. In sum, subjects are able to crossmodally recognize objects that have never crossmodally experienced. Now, by what mechanism can this occur? The next section suggests that *object files* can serve as this mechanism.

5. Object Files as the mechanism for cross-modal perceptual recognition

This section introduces how referential informational content is stored in object files for subsequent recovery when required. Treisman *et al.* (1983) first defined object files as "the temporary representation in which the information that pertains to a particular object accumulates and is updated when the object changes" (p. 531). Since then, the notion has been widely employed in many and varied areas of the psychology and philosophy of perception. Despite a considerable lack of consensus on how object files should be understood, Green and Quilty-Dunn (2021) have recently deployed a thorough characterization.[13] According to these authors, an object file is a perceptual object representation generally characterized as a representation that (i) sustains reference to an external object over time and (ii) stores and updates information concerning the properties of that object (p. 666). The importance of the first point is that the storage of the information is referential and sustained in time. The second point suggests that object files are dynamic and

[13] For very recent criticism about the usability of this notion, see Block (2023) and Nanay (2023). I do not take part in these debates here; I simply assume that Object Files are a useful psychological construct that consistently characterizes how objects are perceptually represented, stored and updated.

may change over time if required by environmental situations. To this general characterization, they add two important points: first, object files represent information in a propositional format, and second, object files store information concerning different feature categories in separate memory stores (p. 666). According to the first point, differences in format between visual and tactile representations (iconic and tactile formats) are irrelevant for subsequent recovery since object files employ a discursive propositional format (i.e., a minimal sufficiently abstracted way to store information). The second point not only assumes that object files are, in essence, memory mechanisms but also that these mechanisms store features of different categories into separate memory stores within an object file.

According to Green and Quilty-Dunn (2021), object files are a construct empirically supported by experimental research. They appeal to three different experimental paradigms. The first are those experiments that examine our ability to maintain and update representations of objects as they move. For example, in the Multiple Object-Tracking (MOT) paradigm, participants should track a subset of flashing items from a bigger set of moving items. Participants are typically effective at tracking up to about 4 items (Pylyshyn 2003) despite significant changes in colour, shape, and size (Zhou *et al.* 2010) or despite intersecting each other or occluding by hidden barriers (Scholl and Pylyshyn, 1999). MOT suggests that there must be a mechanism —object files— that encodes spatiotemporal addresses of objects to maintain identity over time and across changes (Quilty-Dunn, 2020).

The second set of experiments are those that examine our ability to maintain representations of objects after they have disappeared from view; this is the Object-Specific Preview Benefit paradigm (OSPB). In OSPB, subjects are presented with two simple objects with a particular feature for each object (e.g., a letter). At some point, features disappear, and the objects move to new locations. Then, one of the features appears in one object, and subjects should indicate whether it matches one of the original features. Subjects′ performance is faster when the feature re-appears on the same object (compared to when it does on a different object). Kahneman *et al.* (1992) appeal to Object Files to explain this effect. The explanation is that responses are quicker because when the correct feature appears in the appropriate object, subjects benefit from the previewed information, which has been temporarily stored as a stable representation in an object file.

The third is a mechanism called Transsaccadic memory. Transsaccadic memory is the process by which remembered object information is updated across saccades (Irwin & Gordon, 1998; Hollingworth & Henderson, 1998). In order to build a coherent internal representation, the system should represent and update each saccade in memory; i.e., visual constancy, or the experience of the visual scene as continuous, is possible because internal representations are preserved across saccades recorded in visual memory (Higgins & Rayner, 2015). Under this paradigm, subjects are usually instructed to move their eyes toward moving objects and discriminate whether objects change their features or remain unchanged. The studies show that memory-based gaze correction is accurate, fast, automatic, and largely unconscious (Hollingworth *et al.*, 2008; Irwin, 1992; Pollatsek *et al.*, 1984), thus suggesting that representations stored in transsaccadic memory maintain phenomenal coherence across saccades. It is argued that object files modulate this process by maintaining a representation of objects across changes in retinal stimulation caused by eye-movements, i.e., object files enable us to keep the object's identity across saccadic movements (Gordon and Vollmer, 2010; Schut *et al.* 2017; Quilty-Dunn and Green, 2022). Gordon and Vollmer (2010), for example, found a decrease in reaction time for naming a previewed object. This object-specific effect was significantly reduced when typical properties of objects changed, but not when changes did not involve the typical properties of objects —e.g., the correspondence between transsaccadic representations of a banana depending on the yellow colour of the banana, whereas correspondence for transsaccadic representations of an object without a characteristic colour (such as a cube) does not depend on its colour (see also Gordon, 2014). This shows that transsaccadic memory holds the representation from low-level properties and their relation to abstract categories, thus suggesting that abstract information is referentially bonded into object files for subsequent recovery.[14]

[14] For more on the experimental paradigms that suggest the existence of object files, see Quilty-Dunn (2020) and Green and Quilty-Dunn (2021).

Object files act, therefore, as a database that amodally preserves referential perceptual information over time and consists of category-specific slots propositionally stored in memory for subsequent recovery.

6. *Object files and MQ*

The question at hand is, therefore, how object files can help determine a positive answer to MQ. Specifically, we want to know if object files are responsible for recognizing information that has never been seen before. Importantly, it would be inaccurate to say that, according to this approach, the information is literally transferred between sensorial modalities. Instead, the approach suggests that the information is available for any sensorial system capable of interpreting it —there is not, properly speaking, a transference of information from touch to vision, but a visual recognition of referential information previously acquired from touch or any other perceptual system. I argue, then, that object files can be thought of as the mechanisms responsible for storing referential information (regardless of the mode of acquisition), to be recovered in the future, by whatever mode of presentation compatible with the mode of acquisition. Of course, the more presentations, the stronger the imprint, and the easier it will be to recover. The reasons why this appears as a possibility are the properties usually ascribed to object files, properties that are mutually bonded to each other. I focus on four basic properties.

6.1. Object files are referential

This is one of the main characteristics of object files, namely sustaining reference over time (Green and Quilty-Dunn, 2022). As stated above, by referential I understand the connection between the information that impinges the retina or the somatosensorial receptors with a specific conceptual knowledge acquired through learning. There is not much controversy on this point —the above-reviewed literature on the experimental paradigms base their results on the assumption that object files are referential. One way to see this straightforwardly is by considering the difference between perceptual reference and perceptual attribution. We have just seen that referential content may survive some modifications of low-level properties of objects—phenomena like amodal completion (perceptual object recognition even with extensive occluded parts of the objects) or experimental research on multiple object tracking (perceptual object recognition even with changes in low-level features of objects) do not admit discussion on these effects. It seems then that perceptual representation comprises at least two aspects: reference to perceptible individuals and attribution of properties to those individuals. The paradigms described above show that object files are able to maintain the identity of objects even with changes in spatiotemporal, shape, or colour properties. Although these properties are potentially relevant to determining perceptual reference to the object, they are not applicable by virtue of being encoded in the perceptual object representation for that object. Thus, object files are, in this crucial sense, referential.[15]

6.2. Object files are amodal

From the previous property, one can directly uphold that object files must store object representations amodally.[16] Examples of amodal representations are feature-based word meaning representations, semantic networks, schemata, or

[15] There is a very interesting debate on whether perceptual attribution determines or guides perceptual reference (for discussion, see Quilty-Dunn and Green, 2023). Some think that all attribution guides reference (Burge, 2010), others that none attribution guides reference (Pylyshyn, 2009) and others that only a privileged subset of attributions guide reference (Flombaum *et al.*, 2009). This is not relevant for the proposes of this paper, but there are reasons to think that the range of perceptual attributives used to guide reference shifts adaptively with context (Quilty-Dunn and Green, 2023).

[16] For a very recent review of the different uses of the amodal sensory dimension, see Spence and Di Stefano (2024a).

frames of reference. This is central to sustaining the affirmative answer suggested by OFS. MQ subjects not only tactilely perceive the roundness of the sphere or the cube's edges but they also learn to differentiate them conceptually. In this sense, the tactile recognition of the stimuli involves both the contact of the somatosensorial receptors with an external object as well as the referential identification of the object as a distinctive exemplar of a particular category. Once physical contact occurs, the stored referents corresponding to the recognition of a given stimulus connect with the incoming information. Therefore, the internal amodal representations may capture information from one or more modalities, but these representations are, themselves, modality-unspecific, i.e., they constitute an abstract description of what they represent. Object files are, in this crucial sense, modality-independent internal representations of objects.[17] Evidence for this internal amodal object representation has been found in humans (Jordan *et al.*, 2010), mammals (Winters and Reid, 2010), fish (Schumacher *et al.*, 2016) and even in invertebrates such as bees (Solvi *et al.*, 2020). Now, although the amodal structure is manifestly different from the physical structure of the things represented, object files allow the correspondence between external and internal objects; correspondence that can be decoupled from specific sensory modalities. In sum, once the low-level features of an object are integrated, the system stores via object files an abstract amodal representation of this object, which will be available for subsequent recovery; i.e., once amodal storage is supported, the incoming information in one modality may be recognized through another. Therefore, amodality becomes modal at the very moment the senses intervene. That is, when we see an object we have never seen before but about which we have previously obtained information through touch, the amodal information stored with the representational content of that object becomes modal at the very moment that vision intervenes. If this is so, then MQ would move towards a positive response.

6.3. Object files are propositional in format

In strong relation to the previous points, object files store information in a propositional format rather than in an iconic or tactual one. In the preliminaries of this paper, I sustained that any positive answer to MQ must come from the content of the perceptual experience rather than the format in which such content is represented. I think this is clear when we go straight to the experience: if a response to MQ depends on the format in which the information is recorded, then visual and tactual information are depicted in different formats and, therefore, a positive answer to MQ is not possible. For this reason, I argued that any positive answer must involve the content represented. However, the format in which objects are recorded in object files is independent of the format in which it is perceived —objects can be perceived iconically and recorded discursively. Indeed, the referential character of object files and their amodality indicates that they represent their contents in a non-iconic propositional format. This is key for subsequent recovery of the informational content independently of the format in which such content is perceived. Let me expand on this. Formats are general structural properties of representational states that play a role in individuating types of representations. According to Kosslyn *et al.* (2006, p. 8), 'A format is a type of code'. Just as photographs are codified in an iconic format or tactile graphics are codified in a tactile format, sentences are codified in a discursive or propositional one. For example, the sentence 'that is a rose' is structured differently from a picture or a tactile graphic of a rose. All of them share content (rose) but differ in the format in which such content is represented (discursively, visually or tactilely). There are, however, crucial differences between the properties represented in one or the other. Just as perceptual-like representational formats (visual and tactile in this case) can represent low-level properties such as shape, colour or orientation, the properties represented in discursive formats are silent to this information (at least in this case). The discursive represen-

[17] Spence and Di Stefano (2024a, 2024b) have raised doubts on whether shape information is coded amodally by the senses of vision and touch. In the absence of convincing empirical evidence to support the claim of amodal sensory qualities, they suggest restricting the term amodal to abstract cognition rather than sensory perception. This might, in principle, contradict my view, but I think compatibility is possible. Simply, my understanding of object files as a cognitive mechanism that stores amodal and abstract information that mediates, guides and configurates sensory perception, is not at odds with the assertion that shape information is initially coded in a modal way. I am grateful to an anonymous reviewer for bringing this literature to my attention.

tation, on the other hand, provides categorical information (rose) that a layperson in flowers cannot grasp from the perceptual-like representational formats. The idea is that object files are codified discursively, independently of the mode in which the information has been presented (see Quilty-Dunn, 2016). All this is in strong concordance with the positive answer to MQ defended here and straightforwardly connects with the following property of object files.

6.4. Object files are category-specific

We have just seen that object files retain referential information abstractly, propositionally and independently of the modality in which the information has been recorded. A direct consequence is that object files behave as category-specific previews used to determine the referential matching with external objects. A preview benefit, for example, requires that matching feature combinations are ascribed to a specific object category; although for features not initially attributed to the reviewed object category, there will be no benefit (O´Callahan, 2014, p. 11). The idea here is that even with changes in some object properties the system recognizes objects as pertaining to specific categories, i.e., a handful of properties can be diagnostic of the objects category. Green and Quilty-Dunn (2021) argue that feature representations are organized into separate category-specific slots located within an object file. When subjects are asked if they recognize objects (e.g., a sphere), they represent representations of individuals that function computationally as pointers that enable access to these category-specific slots. The case is that this type of storage is not specific to any sense modality. Subjects can store the category 'sphere' with its typical properties into an object file by touch, and when subsequently asked if they are visually facing a sphere they will retrieve the correspondent object file that stores the individual categories of the conceptual references (e.g., sphere) and the feature categories of the spatial properties of spheres (e.g., roundness). According to Quilty-Dunn and Green (2021), this is corroborated by evidence that indicates that "features from the same category compete with one another to a much greater degree than features from different categories, even when the features are bound to the same object"(p. 685). One outstanding consequence of this type of storage is the extremely fast time in which humans categorize objects. Potter *et al.*, (2013), for example, found object recognition as fast as 13ms. from stimulus onset. This extremely fast recognition is a substantial adaptative quality since it allows object recognition without the cost of perceptually codifying too many low-level properties; perhaps only a salient feature is enough for object recognition. This property of object files allows object recognition by retrieving the salient properties stored into the object file that administrates the specific category; again, object files are not modality-specific and therefore nothing prevents this process can being done from whichever sense modality is able to recognize it.

6.5. Taking stock

Once the properties of object files are defined, it is easy to see the mechanism from which a newly sighted subject might recognize tactilely familiar shapes by sight alone. If tactile information is stored in object files in the way defined above, i.e., referentially, amodally, discursively and category-specifically, then this information could be subsequently retrievable throughout the different sense modalities able to decode the type of information deployed. Since shape perception can only be registered visually or tactilely, it will be retrieved from the correspondent object file in one of the two modalities. Thus understood, appealing to object files can offer a positive answer to MQ.

7. Defending OFS

Once OFS is explained, I will provide some empirical research that involves the view. One is the possibility that object files store multimodal information, the other concerns experimental research in robotics and animal experimentation. The former shows that object files store information in a way that transcends sensory modalities, whereas the latter opens up the possibility of connecting empirical research on crossmodal perception with object files, thus providing an avenue towards OFS. Although neither of these is definitive, they both constitute a step towards the OFS's positive response to MQ.

7.1. Multimodal object files

Jordan *et al.* (2010) conducted a study to determine whether the contents of object files are specific to a particular sense (modality-specific) or if they can be accessed across different senses (amodal). The researchers carried out two experiments. The first experiment investigates whether object file correspondence can occur through a combination of visual and auditory information. The second experiment rules out the possibility of explaining the results in the first experiment via verbal encoding. The study indicates that object files are not strictly tied to vision and that object-related information is stored in an amodal format. Therefore, an object file can be initially formed with visual input and later accessed with corresponding auditory information, indicating that object files may operate at a multimodal level of perceptual processing. Researchers conclude that object files "store object-related information in an amodal format that can be flexibly accessed across senses" (p. 500). They leave open the possibility that object files are modality-specific but accessed by any modality able to recover the information, i.e., that there is a cross-modal correspondence, or that object files are devices that store information amodally, in a sort of common code mechanism, in order to be accessed later by any modality whatsoever. Any of these possibilities lean towards an affirmative response to MQ, although the second one is more aligned with the objectives of the OFS.

7.2. Robotics and animal experimentation

As I pointed out above, experimental research on MQ with blindness restoration in humans is inconclusive due to the implausibility of sudden restoration of the visual system. However, alternative approaches involving robotics and non-human animals might provide valuable insights. Studies on robotics (Falco *et al.*, 2019; Liu *et al.*, 2019) and animal experimentation (Versace *et al.*, 2024; Solvi *et al.*, 2020) offer compelling evidence that aligns well with the OFS framework.

Falco *et al.* (2019) trained a classifier by using visual data stored from a camera to recognize objects only with tactile data, without any prior tactile information. This approach represents the opposite scenario from MQ, and as previously mentioned, it is less contentious than the original one. As said, this is likely because of the dominance of the human visual experience over the tactile one, but this dominance can be circumvented by building an artificial system. The point is that the very same mechanism can likely be deployed for both visual-to-tactile and tactile-to-visual transmission of information in humans. These researchers stem from the idea that crossmodal perception necessarily requires common information between the modalities. The way they find this common space is through descriptors. A descriptor is an abstract symbol (words or characteristic features) that collects the attributes of items and stores them in databases to posteriorly enable the system to recognize objects based on data collected (see also Zhang *et al.*, 2016). In this way, the system can recognize objects in any sense modality by matching the information entered from sensors with the descriptors stored in databases. Indeed, although these researchers do not directly refer to an object file mechanism, something similar can be inferred. Likewise, other researchers interested in developing a mechanism to implement an artificial active visual-tactile cross-modal matching designed a shared dictionary learning model which can simultaneously learn from the information coming from the visual and tactile sensors (Liu *et al.*, 2019). All of this is reminiscent in many ways of the referential, amodal, propositional and category-based mechanism proposed here. The crucial point is that to design an artificial mechanism capable of transferring knowledge across various perceptual modalities, researchers must move beyond sensory modalities and leverage algorithmic abstract information.[18]

[18] For a recent overview of visuo-haptic object perception for robots, see Navarro-Guerrero *et al.* (2023).

Of course, one can say that the mechanism generated for object recognition in artificial agents might be far from mirroring the mechanism naturally employed by humans, but in fact, mirroring human behaviour is precisely what researchers in robotics try to do. I have no space here to elaborate further on this issue, but I think that studies in robotics can offer important clues about MQ, mainly because it easily manages to overcome the insurmountable limitations of human experimental situations. Be as it may, it seems that this line of evidence is close to what is suggested in the present paper.

The other line of evidence favouring an affirmative response to MQ consistent with OFS comes from non-human animal experiments. For instance, Versace *et al.* (2024) recently conducted a experiment with chicks. They exposed newly hatched chicks in darkness to either tactile smooth or tactile bumpy stimuli for 24 h, and immediately after they tested them in a visual recognition task. During their first experience with light, chicks exposed to smooth tactile stimuli approached the visual smooth stimulus significantly more than those exposed to bumpy tactile stimuli. It is noteworthy that chicks are precocial; i.e., they hatch with mature visual, proprioceptive and motor systems, so that their perceptive and motor responses can be evaluated already in the first hours of life, thereby avoiding the rapid restoration problem encountered in empirical research involving humans. These findings show that visually inexperienced chicks can solve MQ positively, thus indicating that cross-modal recognition does not require previous multimodal experience. They argue:

> This significant difference in visual preferences in chicks that differed only in tactile experience shows that visually naive chicks learn about objects experienced in the solely tactile modality, and can use representations based on tactile experience to solve a visual recognition task. The ability of newly-hatched chicks to discriminate between visual objects at first sight, based on previous tactile experience, solves Molyneux's problem, showing that cross-modal recognition from tactile to visual sensory modality does not require previous experience with simultaneous multi-modal stimuli. (Versace *et al.*, 2024, p. 4)

The key point is that chicks are capable of spontaneous cross-modal recognition in the absence of previous cross-modal experience. While not explicitly relying on an object file system, the authors predict the existence of an innate system for crossmodal shape recognition. But crucially, this crossmodal system can only be active when chicks first incorporate tactile information (cross-modality is not possible if no stimulus has ever been experienced). This opens the possibility of abstractly storing tactile information into an object file for later visual retrieval. If chicks are indeed born with an innate crossmodal shape system mediated by object files, it is likely to be quite rudimentary but, despite all, such a system enables them to store amodal internal representations of object shape properties.

A similar experiment was conducted to investigate crossmodal perception in bumble bees. Solvi *et al.* (2020) trained bumble bees in the dark to discriminate between spheres and cubes to posteriorly test their visual discriminative ability for the same objects that could only be seen (in the light) through a clear barrier, and not touched. The bees showed a preference for objects that had been previously rewarded in the dark. Conversely, bees trained to visually discriminate the objects spent more time with the objects in the dark (tactile) setting. Researchers conclude their findings as follows:

> Whether bumble bees solve the task by storing internal representations of entire object shapes (cube or sphere) or local object features (curved or flat edge) remains unknown. In either case, our experiments show that bumble bees are capable of recognizing objects across modalities, even though the received sensory inputs are temporally and physically distinct. Bumble bees show a kind of information integration that requires a modality-independent internal representation. This suggests that similar to humans and other large-brained animals, insects integrate information from multiple senses into a complete, globally accessible, gestalt perception of the world around them. (p. 911)

Again, the connection with OFS is clear. Bumble bees retain modality-independent (amodal) internal representations that can be accessed by any perceptual modality capable of decoding them.

This body of evidence supports the OFS approach suggested here. In addition to this, other cross-modal recognition studies in animal cognition, such as cetaceans (Bruce and Pack, 2022) or rats (Reid *et al.*, 2014), as well as cross-modal recognition studies in newborns (Streri, 2012) suggest that the mechanism through which perceptual information is stored and remains accessible for subsequent recovery is an object file mechanism (i.e., a referential, amodal, propositional and category-specific mechanism). It is evident that the complexity of the object file mechanism in adult humans far exceeds the simpler mechanisms found in bees, chicks, and other animals —the rich conceptual repertoire of humans is unparalleled in the animal kingdom. Be that as it may, these studies along with previous research reviewed on robotics, can shed light on the mechanism underlying crossmodal perception and consequently offer valuable clues on how to effectively respond to MQ.

8. Conclusion and future approaches

In this paper, I have argued for a positive answer to MQ based on the Object File Strategy (OFS). Firstly, I discuss the specificities of MQ and argue against its experimental assessment. Then, I focus on crossmodal recognition studies, which demonstrate the interaction between vision and touch in shape perception. However, MQ suggests a different scenario. Despite the inequivocal interaction between the two sensory modalities, the defenders of a negative answer to MQ can claim that if subjects have never experienced visual stimuli, then visuo-tactile cross-modality is not possible. Experiments on crossmodal recognition, indeed, do not consider these effects in blind people.

Once here, I have argued that appealing to OFS can provide insight into these issues. The way and the type of information encoded in object files are crucial to my argument. I argue that if tactile information is encoded referentially and the information encoded is amodal, propositional and category-specific, then the information encoded in object files will be available for any sense modality capable of decoding it and, therefore, a newly sighted might recognize tactilely familiar shapes by sight alone. To support this, I have refered to empirical research. Firstly, I draw on Jordan *et al.* (2010) who demonstrate that object files operate at a multimodal level of perceptual processing, suggesting that object-related information should be stored in an amodal format and therefore it will be flexibly accessed across senses. Secondly, I cite studies on crossmodal recognition in artificial agents and non-human animals, which usually appeal to some sort of common code mechanism. This research not only avoids the experimental pitfalls of human research but also aligns perfectly with OFS.

If all this is on the right track, then an avenue to reconsider a positive answer to MQ opens before our eyes. I urge researchers to delve into MQ by considering the way in which perceptual information is received, stored and retrieved. Since the notion of object file inexorably appears as a crucial element throughout this process, I encourage those interested in this topic to consider these issues from the perspective of OFS.

Finally, I suggest researchers consider OFS through other theories on mind functioning. For instance, OFS is in line with the postulates defended by the Predictive Coding Framework according to which minds predict sensorial information by updating the global world model inserted into our brains (Clark, 2013), and fits perfectly with the Language of Thought hypothesis, the idea that thought is done in a mental language. The Language of Thought hypothesis, first suggested by Fodor (1975), has been currently retrieved by Quilty-Dunn, Porot and Mandelbaum (2022) by extending it to perception. I strongly suggest researchers take these lines of coincidence to put together a mutually consistent theory that englobe all these elements and tries to deploy all of them in a single hypothesis. I prognosticate that such a theory inescapably would offer a positive response to MQ in the terms suggested by the present paper.

Aknowledgements

I am grateful to the Research Group in Analytic Philosophy (Appearances, Singular Content and Fiction), of the Logos Group at the University of Barcelona, for the partial funds they provided me for presenting the results of this research project in the "XI Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España".

REFERENCES

Altieri, N. (2015). Multimodal theories of recognition and their relation to Molyneux's question. *Frontiers in Psychology, 5,* 1547. https://doi.org/10.3389/fpsyg.2014.01547

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience, 4*(3), 324-330. https://doi.org/10.1038/85201

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex, 12*(11), 1202-1212. https://doi.org/10.1093/cercor/12.11.1202

Amedi, A., Raz, N., Azulay, H., Malach, R., & Zohary, E. (2010). Cortical activity during tactile exploration of objects in blind and sighted humans. *Restorative Neurology and Neuroscience, 28*(2), 143-156. https://doi.org/10.3233/RNN-2010-0503

Berkeley, G. (1709/2008). An essay towards a new theory of vision (NTV). In D. Clarke (Ed.), *Philosophical writings* (pp. 1-66). Cambridge University Press. https://doi.org/10.1017/CBO9780511802577.005

Block, N. (2023). Let's get rid of the concept of an object file. In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (Chap. 27). Wiley-Blackwell. https://doi.org/10.1002/9781394259847.ch27

Bruck, J. N., & Pack, A. A. (2022). Understanding across the senses: Cross-modal studies of cognition in cetaceans. *Animal Cognition, 25*(5), 1059-1075. https://doi.org/10.1007/s10071-022-01684-8

Burge, T. (2010). *Origins of objectivity*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199581405.001.0001

Campbell, J. (1996). Molyneux's question. *Philosophical Issues, 7,* 301-318. https://doi.org/10.2307/1522914

Cattaneo, Z., & Vecchi, T. (2011). *Blind vision: The neuroscience of visual impairment*. The MIT Press. https://doi.org/10.7551/mitpress/9780262015035.001.0001

Cheng, T. (2015). Obstacles to testing Molyneux's question empirically. *i-Perception, 6*(4), 1-5. https://doi.org/10.1177/2041669515599330

Cheng, T. (2019). On the very idea of a tactile field, or: A plea for skin space. In T. Cheng, O. Deroy, & C. Spence (Eds.), *Spatial senses: Philosophy of perception in an age of science*. Routledge. https://doi.org/10.4324/9781315146935

Cheselden, W. (1728). An account of some observations made by a young gentleman, who was born blind, or lost his sight so early, that he had no remembrance of ever having seen, and was couch'd between 13 and 14 years of age. *Philosophical Transactions, 402*, 447-450. https://doi.org/10.1098/rstl.1727.0038

Clarke, S. (2015). Investigating what felt shapes look like. *i-Perception, 7*(1), 1-6. https://doi.org/10.1177/2041669515627948

Condillac, E. B. (1754/1930). *Condillac's treatise on the sensations* (G. Carr, Trans.). Favil Press.

Connolly, K. (2013). How to test Molyneux's question empirically. *i-Perception, 4*, 508-510. https://doi.org/10.1068/i0623jc

Degenaar, M. (1996). *Molyneux's problem: Three centuries of discussion on the perception of forms*. Springer. https://doi.org/10.1007/978-0-585-28424-8

Di Stefano, N., & Spence, C. (2024). Perceptual similarity: Insights from crossmodal correspondences. *Review of Philosophy and Psychology, 15*, 997-1026. https://doi.org/10.1007/s13164-023-00692-y

Diderot, D. (1749/1916). Letter on the blind for the use of those who see. In *Diderot's early philosophical works* (pp. 68-141). The Open Court Publishing Company. Retrieved from https://archive.org/details/earlyphilosophic00dide

Evans, G. (1985). Molyneux's question. In A. Noë & E. Thompson (Eds.), *Vision and mind* (pp. 319-349). MIT Press. https://doi.org/10.7551/mitpress/7111.003.0016

Falco, P., Lu, S., Natale, C., *et al.* (2019). A transfer learning approach to cross-modal object recognition: From visual observation to robotic haptic exploration. *IEEE Transactions on Robotics, 35*(4), 987-998. https://doi.org/10.1109/TRO.2019.2914772

Ferretti, G., & Glenney, B. (2021). *Molyneux's question and the history of philosophy*. Routledge. https://doi.org/10.4324/9780429020377

Fine, I., Wade, A. R., Brewer, A. A., May, M. G., Goodman, D. F., Boynton, G. M., Wandell, B. A., & Macleod, D. I. (2003). Long-term deprivation affects visual perception and cortex. *Nature Neuroscience, 6*(9), 915-916. https://doi.org/10.1038/nn1102

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In B. M. Hood & L. R. Santos (Eds.), *The origins of object knowledge* (pp. 135-164). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199216895.003.0006

Fodor, J.A. (1975). *The Language of Thought*. Harvard University Press.

Gordon, R. D. (2014). Saccade latency reveals episodic representation of object color. *Attention, Perception, & Psychophysics, 76*(6), 1765-1777. https://doi.org/10.3758/s13414-014-0683-6

Gordon, R. D., & Vollmer, S. D. (2010). Episodic representation of diagnostic and non-diagnostic object color. *Visual Cognition, 18*(5), 728-750. https://doi.org/10.1080/13506280903004190

Green, E. J. (2022). Representing shape in sight and touch. *Mind and Language, 37*(4), 694-714. https://doi.org/10.1111/mila.12352

Green, E. J., & Quilty-Dunn, J. (2021). What is an object file? *British Journal for the Philosophy of Science, 72*(3). https://doi.org/10.1093/bjps/axx055

Gregory, R. L., & Wallace, J. G. (1963). Recovery from early blindness. In *Experimental Psychology Society Monograph* (Vol. II, pp. 65-129). Heffers. https://digitalcommons.rockefeller.edu/jason-brown-library/92

Held, R. (2009). Visual-haptic mapping and the origin of crossmodal identity. *Optometry & Vision Science, 86*, 595-598. https://doi.org/10.1097/OPX.0b013e3181a72999

Held, R., Ostrovsky, Y., de Gelder, B., Gandhi, T., Ganesh, S., Mathur, U., & Sinha, P. (2011). The newly sighted fail to match seen shape with felt. *Nature Neuroscience, 14*, 551-553. https://doi.org/10.1038/nn.2795

Higgins, E., & Rayner, K. (2015). Transsaccadic processing: Stability, integration, and the potential role of remapping. *Attention, Perception, & Psychophysics, 77*(1), 3-27. https://doi.org/10.3758/s13414-014-0751-y

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General, 137*(1), 163-181. https://doi.org/10.1037/0096-3445.137.1.163

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Frontiers in Psychology, 10*, 2246. https://doi.org/10.3389/fpsyg.2019.02246

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention, and trans-saccadic memory. *Visual Cognition, 5*(1-2), 127-155. https://doi.org/10.1080/713756783

James, T. W., James, K. H., Humphrey, G. K., & Goodale, M. A. (2006). Do visual and tactile object representations share the same neural substrate? In M. A. Heller & S. Ballesteros (Eds.), *Touch and blindness: Psychology and neuroscience* (pp. 139-155). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410615671

Jao, R. J., James, T. W., & James, K. H. (2014). Multisensory convergence of visual and haptic object preference across development. *Neuropsychologia, 56*, 381-392. https://doi.org/10.1016/j.neuropsychologia.2014.02.009

Jordan, K. E., Clark, K., & Mitroff, S. R. (2010). See an object, hear an object file: Object correspondence transcends sensory modality. *Visual Cognition, 18*(4), 492-503. https://doi.org/10.1080/13506280903338911

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175-219. https://doi.org/10.1016/0010-0285(92)90007-O

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195179088.001.0001

Lacey, S., & Sathian, K. (2015). Crossmodal and multisensory interactions between vision and touch. *Scholarpedia, 10*(9), 795. https://doi.org/10.1016/B978-0-12-805408-6.00392-4

La Mettrie, J. (1996). Treatise on the soul. In A. Thomson (Ed.), *La Mettrie: Machine man and other writings* (pp. 41-73). Cambridge University Press. https://doi.org/10.1017/CBO9781139166713.006

Leibniz, G. (1981). *New essays on human understanding* (P. Remnant & J. Bennett, Eds. & Trans.). Cambridge University Press. https://doi.org/10.1017/CBO9781139166874

Levin, J. (2008). Molyneux's question and the individuation of perceptual concepts. *Philosophical Studies, 139*(1), 1-28. https://doi.org/10.1007/s11098-007-9072-5

Levin, J. (2018). Molyneux's question and the amodality of spatial experience. *Inquiry, 61*(5-6), 590-610. https://doi.org/10.1080/0020174X.2017.1372306

Liu, H., Wang, F., Sun, F., *et al.* (2019). Active visual-tactile cross-modal matching. *IEEE Transactions on Cognitive and Developmental Systems, 11*(2), 176-187. https://doi.org/10.1109/TCDS.2018.2819826

Locke, J. (1979). *An essay concerning human understanding* (P. H. Nidditch, Ed.). Clarendon Press. https://doi.org/10.1093/oseo/instance.00018020

Lotze, H. (1887). *Metaphysic* (B. Bosanquet, Ed.). Clarendon Press.

Martin, M. G. F. (1993). Sense modalities and spatial properties. In N. M. Eilan, R. McCarty, & B. Brewer (Eds.), *Spatial representations* (pp. 206-218). Basil Blackwell Ltd. https://doi.org/10.1093/oso/9780198238874.003.0012

Matthen, M., & Cohen, J. (2020). Many Molyneux questions. *Australasian Journal of Philosophy, 98*(1), 47-63. https://doi.org/10.1080/00048402.2019.1603246

Meltzoff, A. N. (1993). Molyneux's babies: Cross-modal perception, imitation, and the mind of the preverbal infant. In N. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 219-235). Basil Blackwell.

Nanay, B. (2020). Molyneux's question and interpersonal variations in multimodal mental imagery among blind subjects. In B. Glenney (Ed.), *Molyneux's question* (pp. 257-263). Routledge. https://doi.org/10.4324/9780429020377

Nanay, B. (2023). Object files. In *Mental imagery: Philosophy, psychology, neuroscience* (online ed.). Oxford University Press. https://academic.oup.com/book/51673

Navarro-Guerrero, N., Toprak, S., Josifovski, J., *et al.* (2023). Visuo-haptic object perception for robots: An overview. *Autonomous Robots, 47*, 377-403. https://doi.org/10.48550/arXiv.2203.11544

Noë, A. (2004). *Action in perception*. MIT Press.

Norman, J. F., Crabtree, C. E., & Norman, H. F. (2006). Aging and the visual, haptic, and cross-modal perception of natural object shape. *Perception, 35*(10), 1383-1395. https://doi.org/10.1038/s41598-022-22141-z

O'Callaghan, C. (2014). Intermodal binding awareness. In D. J. Bennett & C. Hill (Eds.), *Sensory integration and the unity of consciousness*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262027786.003.0004

O'Callaghan, C. (2015). The multisensory character of perception. *Journal of Philosophy, 112*(10), 551-569. https://doi.org/10.5840/jphil20151121035

O'Shaughnessy, B. (1989). The sense of touch. *Australasian Journal of Philosophy, 67*(1), 37-58. https://doi.org/10.1080/00048408912343671

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research, 134*, 427-445. https://doi.org/10.1016/S0079-6123(01)34028-1

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2013). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics, 76*(2), 270-279. https://doi.org/10.3758/s13414-013-0605-z

Prather, S. C., Votaw, J. R., & Sathian, K. (2004). Task-specific recruitment of dorsal and ventral visual areas during tactile perception. *Neuropsychologia, 42*, 1079-1087. https://doi.org/10.1016/j.neuropsychologia.2003.12.013

Prinz, J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/3169.001.0001

Purpura, G., Cioni, G., & Tinelli, F. (2018). Development of visual-haptic transfer for object recognition in typical preschool and school-aged children. *Child Neuropsychology, 24*(5), 657-674. https://doi.org/10.1080/09297049.2017.1316974

Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think.* Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/6137.001.0001

Pylyshyn, Z. (2009). The empirical case for bare demonstratives in vision. In R. J. Stainton & C. Viger (Eds.), *Compositionality, context and semantic values* (Studies in Linguistics and Philosophy, Vol. 85, pp. 215-239). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-8310-5_11

Quilty-Dunn, J. (2016). Iconicity and the format of perception. *Journal of Consciousness Studies, 23*(3-4), 255-263.

Quilty-Dunn, J. (2020). Perceptual pluralism. *Noûs, 54*(4), 807-838. https://doi.org/10.1111/nous.12285

Quilty-Dunn, J., & Green, E. J. (2023). Perceptual attribution and perceptual reference. *Philosophy and Phenomenological Research, 106*(2), 273-298. https://doi.org/10.1111/phpr.12847

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2023). The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences. *Behavioral and Brain Sciences, 46*, e261. https://doi.org/10.1017/S0140525X22002849

Reid, J. M., Jacklin, D. L., & Winters, B. D. (2014). Delineating prefrontal cortex region contributions to crossmodal object recognition in rats. *Cerebral Cortex, 24*(8), 2108-2119. https://doi.org/10.1093/cercor/bht061

Reid, T. (1764/2007). *An inquiry into the human mind on the principles of common sense* (J. Bennett, Ed.). Edinburgh University Press. https://doi.org/10.1515/9781474471923

Rescorla, M. (2020). Perceptual co-reference. *Review of Philosophy and Psychology, 11*(3), 569-589. https://doi.org/10.1007/s13164-018-0411-6

Sathian, K., & Lacey, S. (2022). Cross-modal interactions of the tactile system. *Current Directions in Psychological Science, 31*(5), 411-418. https://doi.org/10.1177/09637214221101877

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38*(2), 259-290. https://doi.org/10.1006/cogp.1998.0698

Schumacher, S., Burt de Perera, T., Thenert, J., & von der Emde, G. (2016). Cross-modal object recognition and dynamic weighting of sensory inputs in a fish. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7638-7643. https://doi.org/10.1073/pnas.1603120113

Schwenkler, J. (2012). On the matching of seen and felt shape by newly sighted subjects. *i-Perception, 3*(3), 186-188. https://doi.org/10.1068/i0525ic

Schwenkler, J. (2013). Do things look the way they feel? *Analysis, 73*(1), 86-96. https://doi.org/10.1093/analys/ans137

Serrahima, C. (2023). Sensory fields: The visual and the bodily. *Philosophical Studies, 180*(2), 679-700. https://doi.org/10.1007/s11098-022-01838-x

Smith, A. (1795). *Essays on philosophical subjects*. Dublin: Printed for Messrs. Wogan, Byrne, J. Moore, Colbert, Rice, W. Jones, Porter, and Folingsby.

Snow, J. C., Strother, L., & Humphreys, G. W. (2014). Haptic shape processing in visual cortex. *Journal of Cognitive Neuroscience, 26*(5), 1154-1167. https://doi.org/10.1162/jocn_a_00548

Solvi, C., Gutiérrez Al-Khudhairy, S., & Chittka, L. (2020). Bumble bees display cross-modal object recognition between visual and tactile senses. *Science, 367*(6480), 910-912. https://doi.org/10.1126/science.aay8064

Spence, C., & Di Stefano, N. (2024). What, if anything, can be considered an amodal sensory dimension? *Psychonomic Bulletin & Review, 31*(5), 1915-1933. https://doi.org/10.3758/s13423-023-02447-3

Spence, C., & Di Stefano, N. (2024). Old and new versions of the Molyneux question: A review of experimental answers. *Philosophy and the Mind Sciences, 5*, Article 11337. https://doi.org/10.33735/phimisci.2024.11337

Streri, A. (2012). Crossmodal interactions in the human newborn: New answers to Molyneux's question. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 88-112). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0004

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*(1), 107-141. https://doi.org/10.1016/0010-0285(82)90006-8

Versace, E., Freeland, L., & Emmerson, M. G. (2024). First-sight recognition of touched objects shows that chicks can solve Molyneux's problem. *Biology Letters, 20*, Article 20240025. https://doi.org/10.1098/rsbl.2024.0025

Winters, B. D., & Reid, J. M. (2010). A distributed cortical representation underlies crossmodal object recognition in rats. *Journal of Neuroscience, 30*(18), 6253-6261. https://doi.org/10.1523/JNEUROSCI.6073-09.2010

Zhang, M. M., Kennedy, M. D., Hsieh, M. A., & Daniilidis, K. (2016). A triangle histogram for object classification by tactile sensing. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4931-4938). https://doi.org/10.1109/IROS.2016.7759724

Zhou, K., Luo, H., Zhou, T., Zhuo, Y., & Chen, L. (2010). Topological change disturbs object continuity in attentive tracking. *Proceedings of the National Academy of Sciences, 107*(49), 21920-21924. https://doi.org/10.1073/pnas.1010919108

Sergio Cermeño-Aínsa holds a PhD in Cognitive Science and Language from the Autonomous University of Barcelona and is a lecturer of Philosophy at the University of Girona (Spain). He has worked mainly on topics belonging to philosophy of mind, philosophy of perception and foundations of cognitive science.

Address: Department of Philosophy, Facultat de Lletres, University of Girona, Edifici Sant Domènec I, Pl. Ferrater i Mora, 1, Campus Barri Vell, Girona 17004, Spain. E-mail: elsergei@hotmail.com - ORCID: https://orcid.org/0000-0002-9598-8386

Theoria, 2026, 41(1), 57-70
https://doi.org/10.1387/theoria.27426

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

Time dilation and rates of the passage of time

(Dilatación temporal y ritmo del paso del tiempo)

Peter J. Riggs*
Australian National University
https://orcid.org/0000-0003-2673-0864

Keywords

Passage of time
Rate of time passing
Time dilation

ABSTRACT: Debate over the issue of the rate of the passage of time has been persisting in the academic literature for decades without substantial progress. The common explanations of the (empirically well-confirmed) time dilation effects from Special and General Relativity theories requires that there is a physical passage of time which occurs at varying rates. Yet, these theories do not formally posit any passage of time. It is shown that the relativistic time dilation effects strongly imply that the passage of time is not a physical phenomenon.

Palabras clave

Paso del tiempo
Ritmo del paso del tiempo
Dilatación temporal

RESUMEN: El debate sobre el ritmo del paso del tiempo ha persistido en la literatura académica durante décadas sin progreso sustantivo. La explicación habitual de los (empíricamente confirmados) efectos de dilatación temporal a partir de las teorías de la relatividad especial y general requiere que haya un paso del tiempo que transcurra a ritmos variables. Sin embargo, estas teorías no postulan formalmente ningún paso del tiempo. En este artículo se muestra que los efectos relativistas de dilatación temporal implican con fuerza que el paso del tiempo no es un fenómeno físico.

Gako-hitzak

Denboraren iragaitea
Denboraren iragaitearen erritmoa
Denboraren dilatazioa

LABURPENA: Hamarkadetan, literatura akademikoan, eztabaidagai izan da denboraren iragaite-erritmoaren auzia aurrerapen nabarmenik gabe. Erlatibitate bereziaren eta orokorraren teoriei jarraikiz (enpirikoki ondo berretsitako) denbora-dilatazioaren efektuak azaltzean, ziurtzat jo ohi da badagoela denboraren iragaite fisiko bat, zeina erritmo aldakorretan gertatzen den. Alabaina, teoria horiek ez dute formalki iradokitzen denboraren inolako iragaiterik badenik. Artikulu honetan erakusten da denbora-dilatazioaren ondorio erlatibistek irmo iradokitzen dutela denboraren iragaitea ez dela fenomeno fisiko bat.

* **Correspondence to:** Peter J. Riggs. Research School of Physics, Mills Road, Australian National University, Canberra ACT 2601 Australia – peter.riggs@anu.edu.au – https://orcid.org/0000-0003-2673-0864

How to cite: Riggs, Peter J. (2026). «Time dilation and rates of the passage of time»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 41(1), 57-70. (https://doi.org/10.1387/theoria.27426).

Received: 19 March, 2025; Final version: 24 October, 2025.
ISSN 0495-4548 - eISSN 2171-679X /

1. Introduction

The idea that time has a "dynamic" nature, i.e. that there is a facet of time characterised by words such as "passage", "flow", or "advance", is very natural for human beings to believe as it seems to be an obvious trait of daily life (Smart 1949, p. 484; Tallant 2016, p. 35; Callender 2017, p. 1). We are all aware, at least intuitively, of this (apparent) "dynamic" aspect of time. However, a perennial puzzle for philosophers who argue in favour of the existence of an objective passage of time (also called temporal passage) is what should constitute its actual description. Various "dynamic" models have been proposed over the last hundred years, e.g. Presentism, Moving Spotlight, Growing Block, etc. (for details see: Dainton 2010; Harrington 2015; Dyke 2021; and Forbes 2024) without any consensus being reached by "dynamic" advocates.

If we do make the general assumption that time does possess a "dynamic" nature then just this conjecture alone brings forth a host of questions, both philosophical and physical. One of the most persistent and troublesome issues about the passage of time is its *rate*. If there is some kind of temporal passage then it stands to reason that there would have to be a rate at which passage occurs (Price 1996, p. 13; Romero 2015, p. 136; Boccardi 2016, p. 9). Most philosophers of time accept that if time passes, it does so at some rate, as Prosser has stated:

> … a significant number of philosophers, both advocates and opponents of [temporal] passage, have felt that it makes sense to speak of time passing at a rate. (Prosser 2013, p. 317)

If time passes then its rate of passage would be an intrinsic trait (i.e. being essential to time's nature) where the value of the rate (i.e. its magnitude) may change depending on physical circumstances (Newman 2021, p. 5). This article will examine how Special and General relativistic time dilation effects might influence time's rate of passage in order to draw a conclusion about the existence of a physical passage of time.

2. Debate over the rate of time's passage

It is the case that deliberations over the issue of the rate of the passage of time have been appearing in the academic literature for more than 80 years. In 1938, C.D. Broad maintained that the rate of time's passage (or the rate of absolute temporal becoming, as he called it) is a primitive concept which cannot be analysed further (Broad 1938, pp. 280-281). Additional discussions followed sporadically throughout the twentieth century (e.g. Smart 1949; Williams 1951; Webb 1960; Prior 1968; Park 1972; Grünbaum 1973; Zwart 1976; Smart 1980; Schlesinger 1982; Kroes 1984; Levison 1987; Markosian 1993; Schlesinger 1994; Dorato 1995; Price 1996) *without agreement* on the issue.

The years since the beginning of the twenty-first century have been witness to greater exploration into (and passionate debate over) definitions of the rate of the passage of time and whether it is a legitimate concept. A survey of the relevant literature verifies this (e.g. see: Lamb 2001; Nerlich 2004; Maudlin 2007; Olson 2009a; Phillips 2009; Raven 2010; Tallant 2010; Price 2011; Skow 2012; Mozersky 2013; Mazzola 2014; Romero 2015; Prosser 2016; Maudlin 2017; Lee 2018; Arthur 2019; Newman 2021; and Gołosz 2022). These arguments will not be revisited. They are cited only to highlight the extent of the continuing debate and its unresolved status.

The unrelenting discussions over the rate of time's passage should not be unexpected as there are many conceptual and empirical difficulties concerning such a rate. The content of the on-going debate over the rate of passage of time shows that questions about time's rate have not been satisfactorily tackled and that there are fiercely held opposing positions. In addition, there has been a tendency of some of the participants in the debate to "talk past each other" (Callender 2017, p. 49). The absence of a shared basis on which to conduct the debate has resulted in there not being any substantial progress achieved.

Whether time does have a "dynamic" aspect or not holds enormous significance for our understanding of human experience and, more generally, the nature of the physical universe. Barry Dainton has described the extent and bearing of the issue as follows:

> The issue of whether or not time is dynamic may impact on how we think of our lives, but it also has consequences of a less parochial sort … *the ontological ramifications* … are vast in comparison. (Dainton 2010, pp. 9-10, italics added)

This significance is a prime motivation for seeking a resolution to the quandary of whether time is or is not "dynamic". An attempt is made below to progress the debate over the passage of time by examining how relativistic time dilation relates to time's rate of passage.

3. Relativistic time dilation and time's passage

In this section, the concepts and equations which are prerequisites for the material offered in Section 4 will be presented. Our two best theories of space and time are Einstein's Special and General Theories of Relativity, both of which are highly confirmed. Neither theory posits any passage of time (Penrose 1989, p. 393). Many philosophers and physicists have argued for over a century that a passage of time does not mesh with Special Relativity and/or is not necessary for a satisfactory account of physical phenomena (e.g. see: Price 1996, pp. 13-15; Mozersky 2000; Dieks 2014; Falk 2016; Turner 2020). Moreover, some commentators forcefully contend that the relativity of simultaneity excludes the possibility of *any* physical passage of time (see: Peterson & Silberstein 2010; Callender 2017, pp. 52-57; Baron 2018). Time is indispensable in the description of the physical world, e.g. the minimum specification of any event requires both space coordinates and a time coordinate. Since time is indispensable, if the passage of time is objective then a parameter for it should be found in physical laws at a basic level. However, there is no term corresponding to any passage of time in the fundamental equations of physics (Morris 1985, p. 209; Greene 2004, p. 130; Callender 2006, p. 498; Al-Khalili 2012, p. 85).

In spite of these arguments and objections, the reality of a passage of time continues to be an intensely held belief with most people, at least in Western countries, accepting that the (apparent) passage of time is a feature of the physical universe (cf. Dainton 2010, p. 28; Prosser 2016, p. 22; Callender 2017, p. 11). Those philosophers who adhere to this belief insist that we only have an *incomplete* description of the universe as the passage of time is absent from physics (see: Dieks 2012, p. 104; Rickles & Kon 2014, p. 3). If this is correct and there is a physical passage of time then temporal passage has eluded mainstream physics.

The Special and General Relativity theories provide descriptions of physical spacetime, i.e. the union of space and time – "flat" spacetime for the former and curved spacetime for the latter (for an intelligible account of spacetime, see Ellis & Williams 2000, pp. 5-11). The metrical structure of spacetime is its most crucial characteristic for the spacetime metric determines intervals and causal relations between events in spacetime (Callender 2017, pp. 122-123). These features of spacetime geometry are (quantitatively) summarised in the spacetime metric tensor (Lawrie 1990, p. 60) which is prominent in the basic equations of Relativity. Yet, as noted above, neither theory explicitly includes temporal passage. The mathematical formulations of Special and General Relativity together with their wide sphere of applications and high degrees of empirical confirmation suggests that these theories still have more to reveal about the physical universe. If the passage of time is an objective feature of the universe then, given the significance of the metric of spacetime, it may well be the case that metrical aspects of time's passage might be *inferred* from an (in-principle) examination of situations where both Special and General relativistic time dilation effects are prominent. We shall investigate this possibility below. Let's first consider these time dilation effects and their equations.

https://doi.org/10.1387/theoria.27426

Suppose a "moving" clock (denoted clock 2) in an inertial frame of reference (i.e. a frame where Newton's First Law holds) measures the time interval between two events to be Δt_2, where the "Δ" symbol indicates an interval. Another clock (denoted clock 1) "at rest" measures the time interval between the same events to be Δt_1. Special relativistic time dilation is the effect when the time interval measured by the "moving" clock is numerically less than the time interval between the same events as measured by the clock "at rest", i.e. $\Delta t_2 < \Delta t_1$. The equation relating these time intervals as measured by the clocks in their respective (inertial) frames of reference is (Faraoni 2013, p. 19):

$$\Delta t_2 = (1 - \mathrm{v}^2 / c^2)^{1/2} \, \Delta t_1 \qquad (1)$$

where v (≥ 0) is the relative speed of the reference frames and c is the speed of light in vacuum (hereafter referred to as light-speed). Note that light-speed has the same (constant) value in all inertial reference frames and is unreachable by any material body, i.e. v is always strictly less than c (Faraoni 2013, p. 36), so that $(\mathrm{v}^2/c^2) < 1$. Relativity textbooks explain why the two time intervals Δt_1 and Δt_2 differ in terms of separate worldlines (i.e. paths in spacetime) between events having unequal time intervals due to the geometrical structure of 'flat' spacetime (Riggs 2022, p. 3).

The speed v is typically the speed of one object with respect to another object. Speed is the magnitude of an object's velocity vector so that the object can have an acceleration without its speed changing by there being only a change in the direction of its motion. Both changes in speed and changes in direction are (formally) accelerations as they alter a velocity vector. Contrary to a widespread misunderstanding, Special Relativity has no difficulty in dealing with accelerated motion (Penrose 2004, p. 422; Arthur 2010, p. 169). The validity of the *Clock Hypothesis* is also accepted which states that the time interval registered on a clock only depends on its speed and not its acceleration (see Arthur 2010 for reasons for acceptance).

It is frequently claimed that the time dilation effect of Special Relativity is explained by there being a physical passage of time which occurs at various rates in different reference frames. This is the common explanation of special relativistic time dilation as found in (print and internet) popularisations of relativity theory. In the common explanation, an actual passage of time with a rate which depends on the relative speed of reference frames is essential. Here is a representative statement of the common explanation of special relativistic time dilation:

> … if you were in a speeding rocket ship, the passage of time inside that rocket would have to slow down with respect to someone on Earth. Time beats at different rates, depending on how fast you move. (Kaku 2008, p. 200)

This slowing of time requires there to be a physical passage of time where its rate has a smaller value than the rate in another reference frame. Moreover, due to the relativity of simultaneity, the physical passage of time would need to be a *local* phenomenon.

Clearly then, in the common explanation, less time would lapse in one frame (e.g. the spacecraft's frame) leading to a shorter time interval between events than in a different frame (e.g. the frame of the Earth). The assumption of a physically slower time thereby accounts for the numerical discrepancy in the time intervals between two events as measured in different reference frames (i.e. (1)). Indeed, it has even been claimed that the slowing of time is the *best explanation* for special relativistic time dilation (Newman 2021, p. 1). However, since relativity theory does not formally express anything about time passing, this common explanation is (in essence) a metaphysical 'add-on'.

In order to make a quantitative analysis of the metrical implications of the passage of time in a relativistic context, we need to employ a frame independent quantity. The apt quantity to use in a *physical* account is proper time (denoted τ) not coordinate time t (Nerlich 2004, p. 23; Peacock 2006, p. 250). A proper time interval is defined between two events which have a timelike separation, i.e. the relativistic spacetime separation applying to material objects, and is reference frame invariant. The proper time interval between two timelike related events (e.g. the reception of two successive pulses of light) can be measured by a clock which shares the worldline joining these events (Adler *et al.* 1975, pp. 122-123; Ferraro 2014, p. 18; Woodhouse 2014, p. 52; Newman 2021, p. 3).

Suppose we have a clock flying past an identical clock which is at rest with respect to a suitable inertial frame with a relative constant speed v close to light-speed. If two successive events which have a timelike separation occur then the time difference between these events measured by the clocks will be their respective proper time intervals, denoted $\Delta\,\tau_2$ for the "moving" clock and $\Delta\,\tau_1$ for the "stationary" clock. We can relate these proper time intervals by replacing the coordinate time intervals $\Delta\,t_1$ and $\Delta\,t_2$ in (1) with $\Delta\,\tau_1$ and $\Delta\,\tau_2$ respectively (Kroes 1985, pp. 78-80; Nerlich 2004, p. 28):

$$\Delta\,\tau_2 = (1 - \mathrm{v}^2 / c^2)^{\,1/2}\,\Delta\,\tau_1 \qquad (2)$$

It must be stressed that proper time intervals are those which have physical significance in both Special and General Relativity.

In General Relativity, gravitational time dilation is where the time interval between two (timelike) events has unequal values at different distances from a source of gravity. A clock further away from the source (i.e. at a "higher" gravitational potential) will record a larger proper time interval than a clock closer to the source. Suppose we have two clocks at rest relative to each other and the gravitational source which are at different distances from the source. Let the radial coordinates of the clocks be r_1 and r_2 from the source of gravity and the proper time intervals registered between two events be $\Delta\,\tau_1$ at r_1 and $\Delta\,\tau_2$ at r_2. The equation relating the proper time intervals measured by these clocks is (Adler *et al.* 1975, p. 136):

$$\Delta\,\tau_2 = [g_{00}\,(r_2) \,/\, g_{00}\,(r_1)]^{1/2}\,\Delta\,\tau_1 \qquad (3)$$

where $g_{00}\,(r_2)$ and $g_{00}\,(r_1)$ are the time components of the spacetime metric tensor (i.e. functions of the gravitational potentials) at positions r_2 and r_1 respectively. If $r_2 > r_1$ then $\Delta\,\tau_2 > \Delta\,\tau_1$. Relativity textbooks explain these unequal proper time intervals in terms of the geometrical structure of curved spacetime (Riggs 2023, p. 3). However, if it is accepted that a physical passage of time occurs at various rates then the explanation of gravitational time dilation is that gravity slows the passage of time as a source of gravity is approached. This is the common explanation of gravitational time dilation which (again) is to be found in print and internet popularisations.

We should also acknowledge that special relativistic and gravitational time dilation effects are so extremely well-confirmed that they must be viewed as being beyond *practical* doubt (see: Hafele & Keating 1972; Bailey *et al.* 1977; Williams 2002, pp. 126-128; NPL 2005; Reinhardt *et al.* 2007; Chou *et al.* 2010; NPL 2011; Bertolami & Páramos 2014; Botermann *et al.* 2014). Such confirmation continues on a daily basis by the operation of the Global Positioning System (see: Ashby 2003; Pascual-Sánchez 2007; Taylor *et al.* 2018, chap.4) It is worth emphasising again that these time dilation effects are thoroughly explained within the two theories of relativity without any need to postulate a physical passage of time.

4. *Quantified rates of time's passage*

If one maintains that the passage of time is physically objective and is a local phenomenon, it would also follow that different rates of passage are local characteristics of time (see: Dieks 2006; Newman 2021). Both special relativistic and gravitational time dilation being localised effects is not incompatible with this conclusion. What might these local rates inform us about the nature of time? In order to pursue this question, it might be expected that a prerequisite is to have a concept of the passage of time that is coherent and a definition of its rate that is consistent. Yet, these are long-standing, unresolved problems in the philosophy of time which remain a primary challenge for passage advocates.

https://doi.org/10.1387/theoria.27426

Progress in the debate over the issue of the passage of time can be achieved by *sidestepping* the definitional problems whilst still making the minimal assumptions that time does have a physical passage and that there are precise (positive, real-valued) rates of passage, regardless of how the rates might be (consistently) defined. Positive, real-valued rates are necessary otherwise the passage of time could either cease or "go backwards". We will also need to specify how rates of passage relate to time intervals between events. This proposed (sidestepping) approach would be compatible with a range of different concepts of time's passage. Since no concept of the passage of time has withstood sustained criticism *and* been found acceptable in both physics and philosophy of time, the proposed approach offers a means to move the debate forward. Let's accept these assumptions and see how far this approach can reach.

We shall begin by (generally) specifying rates of passage of time whilst still not having to define them explicitly. In intergalactic space (where gravitational time dilation is completely negligible), we shall assign the magnitude of the rate of passage to have the value α. Consider the case of a clock (denoted clock 1) which sits in intergalactic space (i.e. zero gravitational time dilation) and is at rest relative to say, the cosmic microwave background (i.e. zero special relativistic time dilation). An identical clock (denoted clock 2) on a spacecraft travelling at constant speed v with respect to clock 1 has previously been synchronised with clock 1. When the spacecraft goes past the position of clock 1, two successive radio or laser pulses are directed at the spacecraft from the location of clock 1 such that the pulses travel the same distance. Let the proper time interval between the pulses recorded on clock 1 be $\Delta \tau_1$ and $\Delta \tau_2$ on clock 2. These proper time intervals are related by (2). The size of the rate of the passage of time at the location of clock 1 has the value α. Given that $\Delta \tau_2 < \Delta \tau_1$, it is the case that time's passage is slowed on the spacecraft and therefore its rate would be reduced by an amount which depends on speed v, as a lower rate yields a shorter time interval and vice-versa. If we denote this variable amount as f then the rate of passage on the spacecraft will have the value $(\alpha - f)$ with $0 \leq f < \alpha$ for $0 \leq v < c$. The quantity f gets larger with increases in speed v such that the rate on the spacecraft tends to zero as its speed advances towards light-speed. The rate cannot be zero as this would require $v = c$.

The common explanation of gravitational time dilation is that the rate of the passage of time *reduces* with decreasing distance to a source of gravity. It would then follow that, as we approach a sizeable gravitational source (e.g. the Earth), time's local passage will slow and its rate will get smaller and smaller. Set the size of the rate at a specified height h above the Earth's surface (where the rate will be determined by the gravitational potential at this height) to have the value $(\alpha - e_2)$ where e_2 is a fixed amount by which time's passage is slowed at height h, with $0 < e_2 < \alpha$. The rate on the Earth's surface will have an even lower value. We can likewise set the size of this rate to be: $(\alpha - e_1)$ where e_1 is a fixed amount by which time's passage is slowed on the Earth's surface, with $e_2 < e_1 < \alpha$. Note that e_1 and e_2 must be strictly less than α or the passage of time would cease.

Let's move to a situation where both special relativistic and gravitational time dilation are concurrent. In this circumstance, we choose to have our spacecraft travelling at a speed v which is a large fraction of light-speed whilst in a circular, equatorial orbit around the Earth so that the value of the gravitational potential for the spacecraft is constant. Such an orbit can be maintained by use of suitably directed thrust from the spacecraft's engines. Clock 1 is on the Earth's surface and clock 2 is on the spacecraft. The equation which relates the proper time intervals between two suitable events for an object travelling in a circular orbit and for an object on the Earth (respectively $\Delta \tau_2$ and $\Delta \tau_1$) has been derived to be (Matolcsi & Matolcsi 2008, pp. 1147-1150):

$$\Delta \tau_2 = [1 + (2\Phi/c^2) - (v^2/c^2)]^{1/2} \Delta \tau_1 \tag{4}$$

where Φ (> 0) is the gravitational potential at the altitude of the orbit. Notice in (4) that the term $[1 + (2\Phi/c^2) - (v^2/c^2)]$ is greater than unity for $(2\Phi/c^2) > (v^2/c^2)$ and less than unity for $(2\Phi/c^2) < (v^2/c^2)$ but is always positive.

We again choose the suitable events to be two successive radio or laser pulses directed at the spacecraft from the location of clock 1 such that the pulses travel the same distance. Note that the proper time interval $\Delta \tau_2$ (measured on the spacecraft) in this situation arises from the competing effects of gravitational and special relativistic time dilation. Gravitational time dilation would lead to a longer time interval between the pulses as measured on the spacecraft than on the Earth's surface and special relativistic time dilation would lead to a shorter time interval. Which effect dominates will depend on the speed of the spacecraft and its altitude (i.e. orbital radius). What might this scenario reveal about time's rate of passage?

Given that the rate at which time passes would have to determine the corresponding time interval between events, there must be a definite quantitative relation which holds between the local rate of time's passage and the proper time interval which results from this passage. What is this relation likely to be? The most straight-forward relation is direct proportionality. If time does pass then there are sound reasons for accepting the relation between time's local rate and the resulting proper time interval to be that of direct proportionality. There are at least three justifications for accepting this relation:

— Harmony with basic natural processes.

The relation of direct proportionality is in harmony with the tendency for all basic natural processes to have the simplest quantitative expression (other things being equal).

— Non-linear changes in the rate of passage are not experienced.

If the human experience of the passage of time is accepted as veridical then the relation of direct proportionality would need to hold as disjointed, discontinuous, or other non-linear changes in the perceived rate of passage are never experienced (by mentally stable people).

— Time shares some characteristics with a smoothly flowing river.

Many philosophers who accept that time passes insist that time shares some (but obviously not all) characteristics with a smoothly flowing river (or similar water course). On this basis, the relevant traits of time would be having a physical passage with a fixed direction and rates that can vary. (A few theoretical physicists also embrace this view of time and use it in concert with the common explanations of time dilation, e.g. Novikov 1998; Kaku 2008). Now it is the case that, in an unrestricted channel of (non-turbulent) flowing water, the amount of water flowing depends directly on the rate of flow (Serway & Jewett 2008, 400). Accepting this as a trait in common with time leads to granting that the passage of time would have a rate which is directly proportional to the resulting proper time interval.

In light of these reasons, it is entirely rational to accept that the relation of direct proportionality holds between the local rate of passage and the resulting proper time interval. If we apply this then we will have the following quantitative relationship for clock 1 (on the Earth's surface):

$$\Delta \tau_1 = \mathrm{K}\,(\alpha - \mathrm{e}_1) \tag{5}$$

where $\Delta \tau_1$ is the proper time interval between the pulses measured on clock 1, K (> 0) is the constant of proportionality whose magnitude is determined by the relativistic spacetime separation between the two pulses and whose units are determined by the units of α (with $\mathrm{e}_1 < \alpha$ or the passage of time would cease). Note that timelike spacetime separations have invariant (positive) values.

If the extent of a proper time interval between two events arises from both special relativistic and gravitational time dilation effects then the net rate of time's passage on the spacecraft will equal the size of the rate at height h minus the variable quantity f, i.e. $(\alpha - \mathrm{e}_2 - f)$ where f depends on speed v of the spacecraft (relative to the Earth). This net rate will be directly proportional to the proper time interval between the two successive pulses as measured on the spacecraft ($\Delta\,\tau_2$), so that for clock 2:

$$\Delta\,\tau_2 = \mathrm{K}\,(\alpha - \mathrm{e}_2 - f) \tag{6}$$

where $f < (\alpha - \mathrm{e}_2)$ or the passage of time would cease. Note that (6) only applies when both time dilation effects are concurrent.

If we square the left-hand and right-hand sides of the equal sign in (4), we get:

$$(\Delta\,\tau_2)^2 = [1 + (2\Phi/c^2) - (\mathrm{v}^2/c^2)]\,(\Delta\,\tau_1)^2 \tag{7}$$

Substituting (5) & (6) into (7) allows a quantitative relationship involving the numerical difference of the squares of the proper time intervals between the two pulses and the numerical difference of the squares of the respective rates of time's passage involved to be derived:

$$(\Delta\,\tau_2)^2 - (\Delta\,\tau_1)^2 = [(2\Phi/c^2) - (\mathrm{v}^2/c^2)]\,(\Delta\,\tau_1)^2 = \mathrm{K}^2\,[(\alpha - \mathrm{e}_2 - f)^2 - (\alpha - \mathrm{e}_1)^2] \tag{8}$$

In (8), we have:

$$(\alpha - \mathrm{e}_2 - f)^2 - (\alpha - \mathrm{e}_1)^2 = (\mathrm{e}_2{}^2 - \mathrm{e}_1{}^2) - 2\alpha(\mathrm{e}_2 - \mathrm{e}_1) - 2f(\alpha - \mathrm{e}_2) + f^2 \tag{9}$$

Although we will avoid explicitly defining rates of passage, we can choose the option of taking rates of passage to be dimensionless. The constant of proportionality K will then have units of time. We may also normalise the rate α to a base value of unity.

Using (8) & (9) and setting the rate α to unity (i.e. $\alpha = 1$), we get the equation:

$$[(2\Phi/c^2) - (\mathrm{v}^2/c^2)]\,(\Delta\,\tau_1)^2 = \mathrm{K}^2\,\{(\mathrm{e}_2 - \mathrm{e}_1)\,[(\mathrm{e}_2 + \mathrm{e}_1) - 2] - 2f(1 - \mathrm{e}_2) + f^2\} \tag{10}$$

Substituting for $(\Delta\,\tau_1)^2$ from (5) into (10) yields:

$$[(2\Phi/c^2) - (\mathrm{v}^2/c^2)]\,\mathrm{K}^2\,(1 - \mathrm{e}_1)^2 = \mathrm{K}^2\,\{(\mathrm{e}_2 - \mathrm{e}_1)\,[(\mathrm{e}_2 + \mathrm{e}_1) - 2] - [2f(1 - \mathrm{e}_2) - f^2]\} \tag{11}$$

Note that e_1 and e_2 have fixed values with $0 < \mathrm{e}_1, \mathrm{e}_2 < 1$ (as $\alpha = 1$) and $\mathrm{e}_1 > \mathrm{e}_2$. It then follows that $(\mathrm{e}_2 - \mathrm{e}_1) < 0$ and $(\mathrm{e}_2 + \mathrm{e}_1) < 2$ so that $(\mathrm{e}_2 - \mathrm{e}_1)\,[(\mathrm{e}_2 + \mathrm{e}_1) - 2] > 0$. Also, as $f < (1 - \mathrm{e}_2)$, $[2f(1 - \mathrm{e}_2) - f^2] > 0$. Since $(2\Phi/c^2) > 0$, $(\mathrm{v}^2/c^2) > 0$, and f is a function of speed v with $f = 0$ when $\mathrm{v} = 0$, it can be seen from (11) that:

$$(2\Phi/c^2) = (\mathrm{e}_2 - \mathrm{e}_1)\,[(\mathrm{e}_2 + \mathrm{e}_1) - 2]\,/\,(1 - \mathrm{e}_1)^2 \tag{12}$$

which is a positive quantity (as required); and

$$2f(1 - \mathrm{e}_2) - f^2 = (1 - \mathrm{e}_1)^2\,(\mathrm{v}^2/c^2) \tag{13}$$

Equation (13) is a quadratic equation in f with the solution:

$$f = (1 - e_2) - [(1 - e_2)^2 - (1 - e_1)^2 (v^2/c^2)]^{1/2} \quad (14)$$

where the minus sign between the two main terms has been chosen to meet the condition that $f = 0$ when $v = 0$. We can see that the variable f is not only dependent on speed v but also on e_1 and e_2. We have arrived at two equations (i.e. (12) & (14)) which can be used to describe time's local rates of passage when both time dilation effects are concurrent. So far – so good!

Using (14), the rate of time's passage on the spacecraft may now be expressed in terms of v, e_1 and e_2:

$$(1 - e_2 - f) = 1 - e_2 - \{(1 - e_2) - [(1 - e_2)^2 - (1 - e_1)^2 (v^2/c^2)]^{1/2}\} = [(1 - e_2)^2 - (1 - e_1)^2 (v^2/c^2)]^{1/2} \quad (15)$$

which (taking the plus square-root) gives a positive value (as required for rates of passage) since $(1 - e_2) > (1 - e_1)$ and $(v^2/c^2) < 1$. We can rearrange (12) to provide the following expression for the quantity $(1 - e_1)^2$:

$$(1 - e_1)^2 = (e_2 - e_1) [(e_2 + e_1) - 2] / (2\Phi/c^2) \quad (16)$$

and by substitution of (16) into (15), we find:

$$(1 - e_2 - f) = \{(1 - e_2)^2 - (v^2/2\Phi) (e_2 - e_1) [(e_2 + e_1) - 2]\}^{1/2} \quad (17)$$

In the circumstances where the spacecraft closely approaches light-speed, the value of $\{(1 - e_2)^2 - (v^2/2\Phi) (e_2 - e_1) [(e_2 + e_1) - 2]\}$ in (17) will become negative as $0 < (1 - e_2)^2 < 1$ and $(v^2/2\Phi)$ will have a very large positive value (e.g. $\approx 10^9$ for high Earth orbit). This will result in the rate of the passage of time on the spacecraft becoming equal to an *imaginary* number as light-speed is approached (i.e. right-hand side of (17)) which *contradicts* the requirement that the rate of time's passage must be positive and real-valued. This is a significant outcome. What then follows? Given that the argument presented is based on there being a physical passage of time with a variable rate (which is an intrinsic attribute), the demonstrated inconsistency implies that the passage of time cannot be an element of physical reality (although time itself, being a part of spacetime, has an objective existence).

5. Physical indicators

It is also appropriate to consider relevant physical indicators in respect to the existence of any physical passage of time as such indicators should have some primacy in the assessment of issues about time:

— The fundamental laws of physics do not contain a term corresponding to a physical passage of time.

If the passage of time was an aspect of physical reality, then there would be a variable term in the fundamental equations of physics for time's passage, which is not the case (as already noted in Section 3).

— Clocks do not measure rates of time passing.

Philosophers who advocate the existence of temporal passage contend that clocks explicitly measure time passing. This assertion is incorrect. What a clock does measure is *intervals of time* (Newton-Smith 1980, p. 156; Kroes 1985, p. 39; Nerlich 2004, p. 24; Olson 2009b, p. 447; Franck 2012, p. 95; Davies 2024, p. 139) —conventionally from midnight for standard clocks or from when started for stopwatches, egg-timers and other timing devices.

— No physical passage of time has been experimentally established.

Any regular, repetitive physical process can be used as a type of clock but, as noted above, clocks do not measure the passage of time. Nor has the passage of time been demonstrated to be a physical quantity which can be measured. Consequently, the passage of time is *not experimentally established*. Indeed, if the passage of time had been so established then we would no longer be debating the issue of whether time passes or not as it would have been shown to be an empirical fact (Riggs 2024, p. 461).

These physical indicators also lead to the conclusion that there is no physical passage of time and so provide additional support for accepting this conclusion.

6. *Awareness of time passing*

The human awareness of the passage of time has always been a compelling reason for believing time's passage to be an objective phenomenon. It is for this reason that the conclusion that there is not a physical passage of time conjures up the obvious question —how is the awareness of passage to be accounted for without invoking temporal passage? Note that the only way that human beings are aware of time passing is through its (apparent) conscious perception (Davies 2002, p. 43; Greene 2004, pp. 139-140; Prosser 2007, p. 77). This being the case, it is *not necessary* for there to be a physical passage of time in order to explain human awareness of time passing as this is explicable by mechanisms which do not require the existence of any (objective) physical passage.

Discussion of the details of these mechanisms of the awareness of passage is beyond the scope of the current article. Indeed, this is an area of on-going research where aspects of physics, philosophy, psychology, and neuroscience intersect and which is still in an early stage of development (e.g. see: Prosser 2016; Riggs 2017; Callender 2017; Gruber *et al.* 2022; Droit-Volet et al. 2023; and Binder 2024).

7. *Final remarks*

The assumptions that time does have a physical passage, that there are precise, positive, real-valued rates of time's passage, and this rate being directly proportional to a resulting proper time interval has led to the conclusion that there is no physical passage of time. In accepting this conclusion, we recall the statement quoted in Section 2 that the ontological ramifications of whether or not time is "dynamic" are vast. A full analysis of these ramifications will have to wait for another occasion. Nevertheless, there are two implications of relevance to issues raised in this article which follow immediately from the conclusion that there is no physical passage of time:

(i) the common explanations of special relativistic and gravitational time dilation are both incorrect; and

(ii) the human awareness of time's passage must originate from processes other than a physical passage of time.

REFERENCES

Adler, R.J., Bazin, M. and Schiffer, M. (1975). *Introduction to general relativity* (2nd ed.). McGraw-Hill.

Al-Khalili, J. (2012). *Black holes, wormholes & time machines* (2nd ed.). CRC Press.

Arthur, R.T.W. (2010). Minkowski's proper time and the status of the clock hypothesis. In V. Petkov (Ed.), *Space, time, and spacetime spacetime: physical and philosophical implications of Minkowski's unification of space and time*. Springer.

Arthur, R.T.W. (2019). *The reality of time flow: local becoming in modern physics*. Springer.

Ashby, N. (2003). Relativity in the global positioning system. *Living Reviews in Relativity* 6, 1–42.

Bailey, J., Borer, K., Combley, F., Drumm, H., Krienen, F. *et al.* (1977). Measurements of relativistic time dilatation for positive and negative muons in a circular orbit. *Nature* 268 (5618), 301-305.

Baron, S. (2018). Time, physics, and philosophy: it's all relative. *Philosophy Compass* 13, e12466.

Bertolami, O. and Páramos, J. (2014). The experimental status of special and general relativity. In A. Ashtekar and V. Petkov (Eds), *Springer handbook of spacetime*. Springer.

Binder, M. (2024). Is the time's flow an illusion? — the issue of the temporality of the conscious experience. *Current Opinion in Behavioral Sciences* 57, 101387.

Boccardi, E. (2016). Recent trends in the philosophy of time: an introduction to time and reality I. *Manuscrito: International Journal of Philosophy* 39 (4), 5-34.

Botermann, B., Bing, D., Geppert, C. *et al.* (2014). Test of time dilation using stored Li+ ions as clocks at relativistic speed. *Physical Review Letters* 113 (12), 120405.

Broad, C.D. (1938). Ostensible temporality. In C.D Broad, *An examination of McTaggart's philosophy* (Vol. II). Cambridge University Press.

Callender, C. (2006). Time in physics. In D.M. Borchert (Ed.), *The encyclopedia of philosophy* (2nd ed.). Thomson Gale.

Callender, C. (2017). *What makes time special?* Oxford University Press.

Chou, C.W., Hume, D.B., Rosenband, T. and Wineland, D.J. (2010). Optical clocks and relativity. *Science* 329 (5999), 1630-1633.

Dainton, B. (2010). *Time and space* (2nd ed.). Acumen.

Davies, P. (2002). That mysterious flow. *Scientific American* 287 (3), 40-44.

Davies, P. (2024). The muddlescape of time. *Timing & Time Perception* 12 (2), 138-143.

Dieks, D. (2006). Becoming, relativity and locality. In D. Dieks (Ed.), *The ontology of spacetime*. Elsevier.

Dieks, D. (2012). The physics and metaphysics of time. *European Journal of Analytic Philosophy* 8 (1), 103-119.

Dieks, D. (2014). Time in special relativity. In A. Ashtekar and V. Petkov (Eds), *Springer handbook of spacetime*. Springer.

Dorato, M. (1995). *Time and reality: space-time physics and the objectivity of temporal becoming*. CLUEB.

Droit-Volet, S., Monier, F. and Martinelli, N.N. (2023). The feeling of the passage of time against the time of the external clock. *Consciousness and Cognition* 113, 103535.

Dyke, H. (2021). *Time*. Cambridge University Press.

Ellis, G.F.R. and Williams, R.M. (2000). *Flat and curved space-times*, 2nd ed. Oxford.

Falk, D. (2016). A debate over the physics of time. *Quanta Magazine* (19 July 2016). URL = https://www.quantamagazine.org/a-debate-over-the-physics-of-time-20160719

Faraoni, V. (2013). *Special relativity*. Springer.

Ferraro, R. (2014). From ather theory to special relativity. In A. Ashtekar and V. Petkov (Eds), *Springer handbook of spacetime*. Springer.

Forbes, G. (2024). *Philosophy of time: the basics*. Routledge.

Franck, G. (2012). How time passes. In R. Buccheri, M. Saniga and M. Stuckey (Eds), *The nature of time: geometry, physics and perception*. Kluwer.

Gołosz, J. (2022). *In defence of a dynamic view of reality*. Jagiellonian University Press.

Greene, B. (2004). *The fabric of the cosmos: space, time and the texture of reality*. Penguin.

Gruber, R.P., Block, R.A. and Montemayor, C. (2022). Physical time within human time. *Frontiers in Psychology* 13: 718505.

Grünbaum, A. (1973). *Philosophical problems of space and time* (*Boston studies in the philosophy of science* volume 12). Reidel.

Hafele, J. and Keating, R. (1972). Around-the-world atomic clocks: observed relativistic time gains. *Science* 177 (4044), 168-170.

Harrington, J. (2015). *Time: a philosophical introduction*. Bloombury.

Kaku, M. (2008). *Physics of the impossible*. DoubleDay Books.

Kroes, P. (1984). Objective versus minddependent theories of time flow. *Synthese* 61, 423-446.

Kroes, P. (1985). *Time: its structure and role in physical theory*. Reidel.

Lamb, A.W. (2001). Temporal dynamics: a phenomenologically based alternative to four-dimensionalist and "point-endurantist" views of time. *The Southern Journal of Philosophy* 39 (2), 235-259.

Lawrie, I.D. (1990). *A unified grand tour of theoretical physics*. Taylor and Francis.

Lee, G. (2018). Explaining away temporal flow – thoughts on Prosser's "experiencing time". *Inquiry* 61 (3), 315-327.

Levison, A.B. (1987). Events and time's flow. *Mind* 96 (383), 341-353.

Markosian, N. (1993). How fast does time pass? *Philosophy and Phenomenological Research* 53, 829-844.

Matolcsi, T. and Matolcsi, M. (2008). Coordinate time and proper time in the GPS. *European Journal of Physics* 29, 1147-1151.

Maudlin, T. (2007). *The metaphysics within physics*. Oxford University Press.

Maudlin, T. (2017). A rate of passage. *Manuscrito: International Journal of Philosophy* 40 (1), 75-79.

Mazzola, C. (2014). Does time flow, at any rate? *Metaphysica* 15 (1), 157-172.

Morris, R. (1985). *Time's arrows: scientific attitudes towards time*. Touchstone.

Mozersky, J.M. (2000). Time, tense, and special relativity. *International Studies in the Philosophy of Science* 14 (3), 221-236.

Mozersky, J.M. (2013). The B-theory in the twentieth century. In H. Dyke and A. Bardon (Eds), *A companion to the philosophy of time*. Wiley-Blackwell.

National Physical Laboratory (NPL) (2005). Demonstrating relativity by flying atomic clocks. *Metromnia* 18, 2-3. URL = http://resource.npl.co.uk/docs/publications/newsletters/metromnia/issue18_einstein.pdf

National Physical Laboratory (NPL) (2011). Time flies. *NPL News*, 12 (1 Feb. 2011). URL = http://web.archive.org/web/20190303092231/https://www.npl.co.uk/news/time-flies

Nerlich, G. (2004). How the twins do it: STR and the clock paradox. *Analysis* 64, 21–29.

Newman, A. (2021). The rates of the passing of time, presentism, and the issue of co-existence in special relativity. *Foundations of Physics* 51 (3), 1-19.

Newton-Smith, W.H. (1980). *The structure of time*. Routledge & Kegan Paul.

Novikov, I. (1998). *The river of time*. Cambridge University Press.

Olson, E. (2009a). The rate of time's passage. *Analysis* 69, 3-9.

Olson, E. (2009b). The passage of time. In R. Le Poidevin, P. Simons, A. McGonigal and R.P. Cameron (Eds), *The Routledge companion to metaphysics*. Routledge.

Park, D. (1972). The myth of the passage of time. In J.T Fraser, F.C. Haber and G.H. Müller (Eds). *The study of time*. Springer.

Pascual-Sánchez, J.-F. (2007). Introducing relativity in global navigation satellite systems. *Annals of Physics (Leipzig)* 16 (4), 258-273.

Peacock, K.A. (2006). Temporal presentness and the dynamics of spacetime. In D. Dieks (Ed.), *The ontology of space-time*. Elsevier.

Penrose, R. (1989). *The emperor's new mind: concerning computers, minds, and the laws of physics*. Vintage Press.

Penrose, R. (2004). *The road to reality: a complete guide to the laws of the universe*. Vintage Press.

Peterson, D. and Silberstein, M. (2010). Relativity of simultaneity and eternalism: in defense of the block universe. In V. Petkov (Ed.), *Space, time, and spacetime: physical and philosophical implications of Minkowski's unification of space and time*. Springer.

Phillips, I. (2009). Rate abuse: a reply to Olson. *Analysis* 69 (3), 503-505.

Price, H. (1996). *Time's arrow and Archimedes' point: new directions for the physics of time*. Oxford University Press.

Price, H. (2011). The flow of time. In C. Callender (Ed.), *The Oxford handbook of philosophy of time*. Oxford University Press.

Prior, A. (1968). *Papers on time and tense*. Oxford University Press.

Prosser, S. (2007). Could we experience the passage of time? *Ratio* 20 (1), 75-90.

Prosser, S. (2013). The passage of time. In H. Dyke and A. Bardon (Eds), *A companion to the philosophy of time*. Wiley-Blackwell.

Prosser, S. (2016). *Experiencing time*. Oxford University Press.

Raven, M. (2010). Can time pass at the rate of 1 second per second? *Australasian Journal of Philosophy* 88, 1-7.

Reinhardt, S., Saathoff, G., Buhr, H., Carlson, L.A., Wolf, A. *et al.* (2007). Test of relativistic time dilation with fast optical atomic clocks at different velocities. *Nature Physics* 3, 861-864.

Rickles, D. and Kon, M. (2014). Interdisciplinary perspectives on the flow of time. *Annals of the New York Academy of Sciences* 1326, 1-8.

Riggs, P.J. (2017). The perceptions and experience of the "passage" of time. *The Philosophical Forum* 48 (1), 3-30.

Riggs, P.J. (2022). Four-velocity and time in special relativity. *The Physics Educator* 4 (4), 2250017.

Riggs, P.J. (2023). Free fall motion and time in general relativity. *The Physics Educator* 5 (2), 2350006.

Riggs, P.J. (2024). Questioning the emergence of time. *Journal for General Philosophy of Science* 55 (3), 459-468.

Romero, G.E. (2015). Present time. *Foundations of Science* 20, 135-145.

Schlesinger, G.N. (1982). How time flies. *Mind* 91, 501-523.

Schlesinger, G.N. (1994). *Timely topics*. St. Martin's Press.

Serway, R.A. and Jewett Jr., J.W. (2008). *Physics for scientists and engineers* (7th ed.) Brooks Cole.

Skow, B. (2012). One second per second. *Philosophy and Phenomenological Research* 85 (2), 377-389.

Smart, J.J.C. (1949). The river of time. *Mind* **58**, 483-494.

Smart, J.J.C. (1980). Time and becoming. In P. van Inwagen (Ed.), *Time and cause: essays in honor of Richard Taylor*. Reidel.

Tallant, J. (2010). A Sketch of a Presentist Theory of Passage. *Erkenntnis* 73, 133-140.

Tallant, J. (2016). Temporal passage and the "no alternate possibilities argument". *Manuscrito: International Journal of Philosophy* 39 (4), 35-47.

Taylor, E.T., Wheeler, J.A. and Bertschinger, E. (2018). *Exploring black holes: introduction to general relativity*. Addison-Wesley-Longman.

Turner, J. (2020). Why special relativity is a problem for the A-theory. *The philosophical quarterly* 70 (279), 385-406.

Webb, C.W. (1960). Could time flow, if so how fast? *Journal of Philosophy* 57, 357-365.

Williams, D.C. (1951). The myth of passage. *Journal of Philosophy* 48, 457-472.

Williams, W.S.C. (2002). *Introducing special relativity*. Taylor and Francis.

Woodhouse, N.M.J. (2014). Relativity today. In A. Ashtekar and V. Petkov (Eds), *Springer handbook of spacetime*. Springer.

Zwart, P.J. (1976). *About time: a philosophical inquiry into the origin and nature of time*. North-Holland.

Peter J Riggs is a physicist and philosopher of science at the Australian National University. His current research interests include: the foundations of physics; the philosophy of space and time; and methodology of science.

Address: Research School of Physics, Mills Road, Australian National University, Canberra ACT 2601 Australia. E-mail: peter.riggs@anu.edu.au – ORCID: https://orcid.org/0000-0003-2673-0864

Theoria, 2026, 41(1), 71-86
https://doi.org/10.1387/theoria.27140

THEORIA

ISSN 0495-4548 – eISSN 2171-679X

A critical analysis of cosmological typicality and the anthropic principle

(Análisis crítico de la tipicidad cosmológica y del principio antrópico)

Eduard Dàvila*
University of Valladolid
https://orcid.org/0000-0001-5419-8591

Keywords

Anthropic Principle
Mediocrity Principle
Typicality
Reference class

ABSTRACT: Stephen Weinberg's prediction of the cosmological constant (Λ) represented an influential application of the Anthropic Principle in cosmology. His approach relied on key assumptions including: a multiverse framework, a proportionality between the number of observers and the formation of galaxies, uniform priors for the vacuum energy density within the anthropic range, and the assumption of typicality. While Weinberg's model successfully constrained Λ within observationally reasonable limits, its methodological foundations may raise some concerns. This paper examines three central components of Weinberg's reasoning: the choice of probability measure, the conditionalization scheme linking observer abundance to galaxy formation, and the assumption of typicality. Particular attention is given to the role of self-locating uncertainty in translating observer counts into predictive weight. While Weinberg's framework is internally coherent once its assumptions are granted, I argue that its predictive force depends on substantive commitments concerning prior distributions and the resolution of indexical uncertainty. Recognizing the conditional structure of these assumptions clarifies the epistemic status of anthropic predictions and highlights the need for greater methodological precision in future cosmological applications of observer-based probability.

Palabras clave

Principio Antrópico
Principio de Mediocridad
Tipicidad
Clase de referencia

RESUMEN: La predicción de la constante cosmológica (Λ) realizada por Stephen Weinberg constituyó una influyente aplicación del principio antrópico en cosmología. Su enfoque se basaba en varios supuestos clave, entre ellos: un marco de multiverso, una proporcionalidad entre el número de observadores y la formación de galaxias, «priors» uniformes para la densidad de energía del vacío dentro del rango antrópico, y la suposición de tipicidad. Aunque el modelo de Weinberg logró restringir Λ dentro de límites observacionalmente razonables, sus fundamentos metodológicos pueden suscitar algunas dudas. Este artículo examina tres componentes centrales del razonamiento de Weinberg: la elección de la medida de probabilidad, el esquema de condicionalización que vincula la abundancia de observadores con la formación de galaxias, y la suposición de tipicidad. Se presta especial atención al papel de la incertidumbre auto-localizadora en la traducción del número de observadores en peso predictivo. Aunque el marco de Weinberg es internamente coherente una vez que se conceden sus supuestos, sostengo que su poder predictivo depende de compromisos sustantivos relativos a las distribuciones a priori y a la resolución de la incertidumbre indexical. Reconocer la estructura condicional de estos supuestos permite aclarar el estatus epistémico de las predicciones antrópicas y pone de relieve la necesidad de una mayor precisión metodológica en futuras aplicaciones cosmológicas de probabilidades basadas en observadores.

* ***Correspondence to:*** Eduard Dàvila. Department of Philosophy, University of Valladolid, Plaza Campus Universitario, s/n, 47011 Valladolid, Spain – eduard.davila@estudiantes.uva.es – https://orcid.org/0000-0001-5419-8591

How to cite: Dàvila, Eduard (2026). «A critical analysis of cosmological typicality and the anthropic principle»; *Theoria. An International Journal for Theory, History and Foundations of Science*, 41(1), 71-86. (https://doi.org/10.1387/theoria.27140).

Received: 30 November, 2024; Final version: 13 March, 2026.
ISSN 0495-4548 - eISSN 2171-679X / © 2026 UPV/EHU Press

Gako-hitzak

Printzipio antropikoa
Mediokritatearen printzipioa
Tipikotasuna
Erreferentzia-klasea

LABURPENA: Stephen Weinberg-ek konstante kosmologikoa (Λ) aurresan izana printzipio antropikoaren aplikazio garrantzitsu bat izan zen kosmologiaren arloan. Haren ikuspegia funtsezko zenbait suposiziotan oinarritzen zen, besteak beste hauetan: esparru multibertsal bat, behatzaile kopuruaren eta galaxien eraketaren arteko proportzionaltasuna, hutsaren energia-dentsitaterako «prior» uniformeak tarte antropikoaren barruan, eta tipikotasunaren suposizioa. Weinberg-en ereduak egokiro lortu zuen Λ zedarritzea behaketan arrazoizkoak diren mugen barruan, baina eredu horren oinarri metodologikoek zenbait zalantza eragin ditzakete. Artikulu honek Weinberg-en arrazoibidearen hiru osagai nagusi aztertzen ditu: probabilitatea zenbatesteko egindako aukeraketa, behatzaileen ugaritasuna galaxien eraketarekin lotzen duen baldintzapen-eskema eta tipikotasunaren suposizioa. Arreta berezia jartzen zaio autolokalizazioaren ziurgabetasunak behatzaileen kopurua aurresan-pisu bihurtzean betetzen duen rolari. Nahiz eta Weinberg-en markoak, haren suposizioak ontzat hartuz gero, barne-koherentzia osoa duen, argudiatzen dut haren aurresan-ahalmena *a priori*ko banaketei eta ziurgabetasun indexikalaren ebazpenari buruzko funtsezko konpromisoen mende dagoela. Behin suposizio horiek baldintzazko egitura dutela aitortutakoan, argi gelditzen da aurresan antropikoek estatus epistemikoa dutela, eta agerian uzten da zehaztasun metodologiko handiagoa behar dela behatzaileetan oinarritutako probabilitatea etorkizuneko aplikazio kosmologikoetan erabiltzeko.

1. Introduction

The philosophy of cosmology explores fundamental questions about the universe, such as its origin, evolution, but also the implications of our existence as observers. And every time a cosmological theory is developed, a major challenge is faced: the scarcity of direct observational data. This forces reliance on theoretical principles and probabilistic reasoning to establish a conceptual framework within which well-founded hypotheses can be formulated.

One common strategy to address this type of reasoning is the application of the Anthropic Principle (AP), a topic of extensive discourse (Barrow & Tipler, 1986; Bostrom, 2002a; Carter, 1974). This principle attempts to explain the universe we observe by examining the prerequisites for our existence and how they influence our interpretations. The AP is a self-referential constraint that acknowledges the inherent bias introduced by our mere existence, as it determines observable physical laws and cosmological parameters. Accordingly, our knowledge of the universe is inevitably filtered through the lens of our own existence because every alternative scenario that do not permit observers like us remains unobservable.

Stephen Weinberg (1987, 1989, 1996, 2000) proposed an influential application of the AP to address the long-standing cosmological constant (Λ) problem, which was the following. Although the exact value of Λ was unknown in the mid-1980s, a discrepancy with the calculated magnitude had been partially identified and an explanation was demanded. Weinberg theorized that Λ had to be constrained by the conditions necessary for the existence of galaxies — thus the potential to host observers. Since a sufficiently large Λ would prevent the formation of gravitationally bound structures, Weinberg argued that anthropic selection effects could provide a natural explanation for the observation of a relatively small —but nonzero— cosmological constant.

The problem was indeed severe. Once Λ was measured, it revealed a dramatic discrepancy of approximately 120 orders of magnitude between the predicted value by the standard quantum field theory and the observed values (10^{112} erg/cm^3 vs 10^{-8} erg/cm^3) (Carroll, 2004). However, in contrast to this failure of conventional physics, Weinberg's anthropic calculations had successfully predicted a value of Λ in reasonable agreement with observations. This remarkable achievement demonstrated the potential of anthropic reasoning to constrain fundamental physical parameters, even in the absence of a complete theoretical mechanism to determine Λ from first principles (Polchinski, 2005).

Actually, Weinberg did not aim to uncover the mechanism that determined the cosmological constant, nor its probability within the distribution of possible values. Instead, he demonstrated the likelihood that a given value would be observable, using the formation of galaxies as the only major restriction to establish an upper limit to the universe's vacuum energy, from which Λ was derived. He also assumed the following assumptions:

— The existence of a multiverse (H1): different regions of the multiverse display particular values of Λ, and our observable universe is just one realization of these possibilities.

— Observer dependence on galaxy formation (H2): the number of observers is proportional to the number of galaxies, because gravitational condensation of matter is the primary factor in the emergence of life.

— Uniform priors on the anthropic range (H3): a uniform distribution of the values of Λ characterizes the narrow anthropic range where life-permitting conditions hold.

— The assumption of typicality (H4): we should consider ourselves typical among all possible observers in the multiverse.

Weinberg's assumptions were elegant in their simplicity but relied on methodological choices that called into question certain aspects of their validity. As Azhar (2014, 2015, 2016, 2017) has pointed out, three major challenges emerge when attempting to formalize any cosmological theory: the measurement problem, the framework for conditionalization, and the typicality assumption. And each of Azhar's challenges offers a basis to examine Weinberg's assumptions and evaluate their coherence and justification.

The first issue concerns the definition of a probability measure to count universes, a crucial step in justifying why certain Λ values should be deemed more likely than others. In general, resolving the measurement problem in cosmology is far from trivial because in a vast and potentially infinite universe, where every possible observable event could occur countless times, the calculation of a certain proportion may be impossible if infinities appear in the quotients. Even though different tactics have been described to overcome this issue (Freivogel, 2011; Page, 2008; Vilenkin & Yamada, 2020), choosing the right measurement is generally a controversial aspect. Both the definition of the sample space and the calculation of the probabilities of the different elements that compose it are difficult.

In an effort to resolve the measurement problem for Λ, Weinberg assumed that the distribution of values across the multiverse followed a well-defined probability distribution, at least within the observable range —the only region of interest. Assumptions H1 & H3 enabled him to make a statistical prediction about what observers should expect to find. But Weinberg's choice of a uniform prior over the anthropic range of Λ seemed to introduce an additional layer of arbitrariness. The assumption that all life-permitting values were equally probable may lack a clear physical or epistemic justification, and alternative priors could lead to quite different conclusions. This raises the first important question: to what extent did Weinberg's result depend on an unjustified assumption about prior probabilities, rather than a genuine anthropic necessity?

Once one accepts that a probability measure can be meaningfully defined, the next challenge deals with how probabilities should be updated in light of the observer selection effects. Weinberg's H2 assumption relied on conditioning the probability distribution of Λ on the fact that we observe a life-permitting universe. Thus, our prior knowledge had to be refined by including only those universes capable of supporting observers, and Weinberg chose the formation of galaxies as the way to account for it. In essence, the number of observers in any universe was assumed to scale with the quantity of matter that condensed into galaxies, which was itself determined by mass conservation and Friedman equations. Then, by imposing constraints on the early universe density perturbations and the vacuum energy density, the same equations allowed him to derive an upper threshold for the cosmological constant. However, this conditionalization step was far from trivial. The methodological question of whether observer selection effects genuinely constrain physical parameters, like Λ, remains contentious, as different prior probability distributions could lead to vastly different posteriors. If the anthropic range of Λ was itself an artifact due to how the priors were defined, then the problem of conditionalization may reduce to an issue of priors dependence rather than an independent validation of the argument. Being so, did Weinberg's approach truly reveal why our universe has the observed value of Λ, or did it simply restate an assumption in a way that appeared explanatory?

A final and more subtle difficulty emerges in how Weinberg defined a typical observer in the multiverse. On the one hand, assumption H2 established that the number of observers was proportional to the number of galaxies, implying that universes with more galaxies should have greater weight. Consequently, this assumption played a crucial role in establishing the expectation value of Λ, as it dictated how universes had to be counted in anthropic reasoning. However, his approach also presupposed that galaxy formation was an adequate proxy for intelligent life, an assumption that may not hold. If the likelihood of the emergence of observers varied significantly by other factors, then weighing universes uniquely by their galaxy count would introduce a potentially misleading bias. Moreover, if the weight of each universe was exclusively determined by the number of observers it hosted —without considering any additional property that governed the probability of the universe appearing in the first place— then the outcome becomes trivial: we should always expect to find ourselves in the universe that contains the most observers. Under this assumption, anthropic selection simply directed us to the most populated universe, where it was assumed we were typical, without providing any deeper insight into the fundamental distribution of Λ. And this raises another concern: did such a distribution even impose a meaningful threshold for applying anthropic reasoning?

A deeper issue also arises from assumption H4, concerning our declared typicality. Weinberg implicitly assumed the existence of a well-defined reference class of observers (sufficiently similar to us), but this was nontrivial because the probability distribution of universes was itself weighted by the number of observers they contained. So, if the number of observers defined the distribution, then the very act of constructing a probability measure depended on how the reference class was chosen in the first place. This may introduce a self-referential element: we indeed may assume that we are typical among a group of observers, but the definition of that group in itself shapes the probability distribution that we seek to derive. What would happen to the statistical weighting if the reference class was broader or narrower than expected? Did Weinberg's assumption subtly impose constraints on the distribution of observer that were not physically justified? If the reference class was indeterminate, then the statistical reasoning used to derive Λ may not be as robust as it initially appeared.

The goal of this paper is to critically examine the assumptions of a case study where the Anthropic Principle was successfully applied to develop a cosmological model. Weinberg's prediction of the cosmological constant will be analysed through the lens of Azhar's framework, with particular attention to whether assumptions such as typicality and observer selection introduce elements of arbitrariness. Section 2.1 will examine the measurement problem, focusing on the assumption of equiprobability within the anthropic observational range. Section 2.2 will address the conditionalization schema, which relies on the proportionality between the number of observers and the number of galaxies. Section 2.3 will investigate the implications of considering humanity as typical. After this analysis, Section 3 will propose an alternative approach that may provide a new foundation for future studies.

2. *Critical examination of the framework*

Weinberg's approach to predicting the cosmological constant through anthropic reasoning represented a very influential inference under uncertainty in cosmology studies. His framework, grounded in the idea that our observed Λ is neither arbitrary nor unnatural but constrained by observer selection effects, provided a compelling alternative to fine-tuning arguments. However, as with any theoretical model incorporating anthropic principles, its success depended not only on its numerical predictions but also on the soundness of its methodological assumptions. Notably, Bousso and collaborators (2007) later demonstrated, using a method that relied on less speculative assumptions, that Weinberg's prediction fell short three orders of magnitude. This deviation suggests that reconsidering the original hypotheses may provide a more accurate framework.

A careful examination reveals that Weinberg's approach, while pragmatic and successful, relied on simplifying assumptions that may have introduced an arbitrariness that affected its epistemic robustness. The upcoming sections will analyse Weinberg's reasoning through the lens of the three major conceptual challenges identified by Azhar: the problems of measurement, conditionalization, and typicality. These challenges do not merely expose technical difficulties but rather fundamental issues regarding how probabilities were assigned, updated, and interpreted in a multiverse scenario.

They call into question whether Weinberg's predictions were the result of a genuine anthropic necessity or if they came closer to an artifact of model-dependent choices.

2.1. Confronting H1 & H3: issues on the measurement

By employing a probability updating scheme grounded in the anthropic observational bias, Weinberg achieved a remarkable approximation. The anthropic constraint offered an explanation on the scale of 20σ if the whole sample space of the vacuum energy density (*pv*)—from which Λ could be derived—was considered. Indeed, his remarkable success clarified what was 'observable' versus what was 'possible'. Yet, if the values of *pv* were uniquely evaluated within the anthropic observable range, then Weinberg's accuracy remained in a much lower percentile. The success of his prediction is moderate when it is referred only to what we could actually measure.

The Anthropic Principle was effectively used as a starting point to impose certain limits on the possible values of *pv*. Nonetheless, once the principle was embraced, the distribution of the remaining values was considered equiprobable, thus the idea of further observer selection effects was implicitly abandoned. With regards to this possibility, it is important to acknowledge that models grounded in theoretical frameworks can only be accurately constructed in the absence of inherent biases (think of the Malmquist bias in astronomy).[1] As it has been affirmed, cosmological theories can be highly speculative, and they may risk overlooking unnoticed biases (Ellis, 2007) and this may be one example of it.

Before addressing the issue of assuming an equiprobable distribution of *pv*, it is important to note that Weinberg incorporated an additional stipulation: the range of anthropically allowed values of the vacuum energy density was so narrow, that all other fundamental constants of the universes could be considered as fixed. However, as demonstrated elsewhere (Tegmark *et al.*, 2006), severe variations in these constants can result in universes that appear identical to our own despite having different values of *pv*. As it was shown, assuming uniform priors for *pv* in predictive models to obtain universes similar to ours did not necessarily correspond to uniform priors for derived expressions of *pv* with other fundamental constants.[2] Also, universes like ours could be largely obtained with lower values of *pv* if other fundamental parameters were allowed to vary (even by several orders of magnitude), so a simpler model with fixed parameters and uniform priors may tend to overestimate *pv* towards higher values. But even though Weinberg's model may not fully capture the complexity of the universe's dynamics, particularly given the challenge of incorporating other fundamental constants whose distributions are even less understood than that of *pv*, it is worth acknowledging his audacity of developing a simplified model that brought the theoretical prediction closer to observation.

Weinberg co-formulated the probability distribution that a random observer in any subuniverse could measure values of *pv* in a given range (Martel *et al.*, 1998, p. 4, eq. 1). According to Bayesian statistics, $P_{obs}(pv)$ corresponded to the mean number of observers in universes with a specific *pv* divided by the total observations made in the full set of observable universes, that is,

$$P_{obs}(pv) = \frac{N(pv)\times P(pv)}{\int_0^\infty N(pv)\times P(pv)\partial pv} \qquad \text{eq. 1}$$

[1] This effect occurs when observing cosmological entities (galaxies, quasars, stars, etc.) and an increase in the average luminosity with distance is detected. This simply occurs because fainter sources are no longer detected as distance increases.

[2] i.e. pv/ξ^4Q^3, which relates the vacuum energy density with the matter per photon and the scalar fluctuation amplitude. If a predictive model to obtain universes like ours is based on this expression, the value of *pv* loses power relative to other fundamental parameters with selection effects against galaxy-forming dense halos. This could mean that other fundamental constants may have greater significance and should not be dismissed too quickly.

https://doi.org/10.1387/theoria.27140

Since one did not know the distribution of observers in each universe with a specific *pv*, nor the distribution of *pv* itself, eq. 1 seemed irresolvable. In order to find a solution, Weinberg stated that "the range of anthropically allowed values of *pv* is so much smaller than the energy densities typical of elementary particle physics that, within this narrow range, we can take the *a priori* probability distribution P(*pv*) to be constant" (Martel *et al.*, 1998, p. 5, eq. 2). As P(*pv*) was equalized, the term could be eliminated from both the numerator and the denominator. At the same time, as N(*pv*) had been previously related to galaxy-forming matter thanks to assumption H2 (see section 2.2 for further details), it could then be expressed as a function of known physics equations that described how matter condensed to form galaxies, and $P_{obs}(pv)$ could be finally calculated.

However, there is a problem with removing P(*pv*) from eq. 1 and leaving the weight of observing a certain *pv* to be exclusively determined by N, the number of observers. This simplification, though effective as it obtained a successful prediction, relied on an implicit assumption that was not entirely justified. By eliminating an initially unknown distribution of *pv* only by presuming it to be uniform, the model converted an epistemic limitation into a tractable predictive structure. One could ask whether the method genuinely constrained *pv* or merely concealed the lack of prior knowledge behind a *post hoc* statistical justification. Specifically, there appears to be a problematic shift where neutrality about a distribution (ignorance) was misinterpreted as an argument for low probability (knowledge), an unjustified inference that has been critiqued in other contexts (Benétreau-Dupin, 2015; Norton, 2010). This plausible fallacy will be further discussed in Section 2.3.

The premise of equiprobability, as stated in H2, emerges as fundamentally problematic when scrutinized. First, from a mathematical perspective, assuming a uniform distribution for *pv* may be misleading, as multiple studies suggest that the underlying probability distribution is not flat at all (Garriga & Vilenkin, 2000). In the inflationary framework employed by Weinberg, the values of *pv* were contingent upon the scalar field potential, meaning that the natural variation of *pv* across the multiverse followed a distribution that could deviate significantly from uniformity. Then, if the probability density of *pv* was inherently skewed, Weinberg's assumption of equiprobability within the anthropic range was not derived from fundamental physics but rather imposed for simplicity. Again, this technique proved to be useful, but at the same time it also casts doubt on whether the predicted value of *pv* was explained by anthropic selection or if it was merely an outcome of an arbitrary prior assumption.

Second, from a formal perspective, the appearance of uniformity in the probability of *pv* may be an artifact of scale-dependent comparisons between the anthropic and non-anthropic domains (see fig. 1). When the full multiverse is considered —including regions where *pv* is so large that no galaxies form— the anthropic range appears compressed within this much broader distribution. This can create the illusion of constancy, much like the size differences between microscopic and macroscopic living organisms would seem negligible when plotted against the vast scale of galaxies.[3] However, once selection effects come into play and our attention is restricted only to the anthropically viable range, its probability density is no longer seen in the context of the full distribution. In this restricted domain, significant variability may emerge, and the probability function may display extreme differences that were previously masked by scale effects. Thus, Weinberg's conjecture only holds if the entire range of possible values is considered in his probability assignments. But, if one removes the broader set and retains only the anthropic subset, it is no longer methodologically sound to assume that properties derived from the full range, such as uniformity, still apply.

---

[3] Consider a hypothetical plot displaying the sizes of various entities in the universe, ranging from subatomic particles to entire galaxies. Within such a scale, the region representing living organisms would appear highly compressed. One might then claim that "all living creatures are roughly the same size", as their variations seem negligible compared to astronomical scales. If this observation were used to define a universal measuring scale, and attention were later restricted to biological sizes using that same ruler, the result would be misleading: subtle variations would be obscured, and extreme cases —such as microscopic life— might go entirely unnoticed.

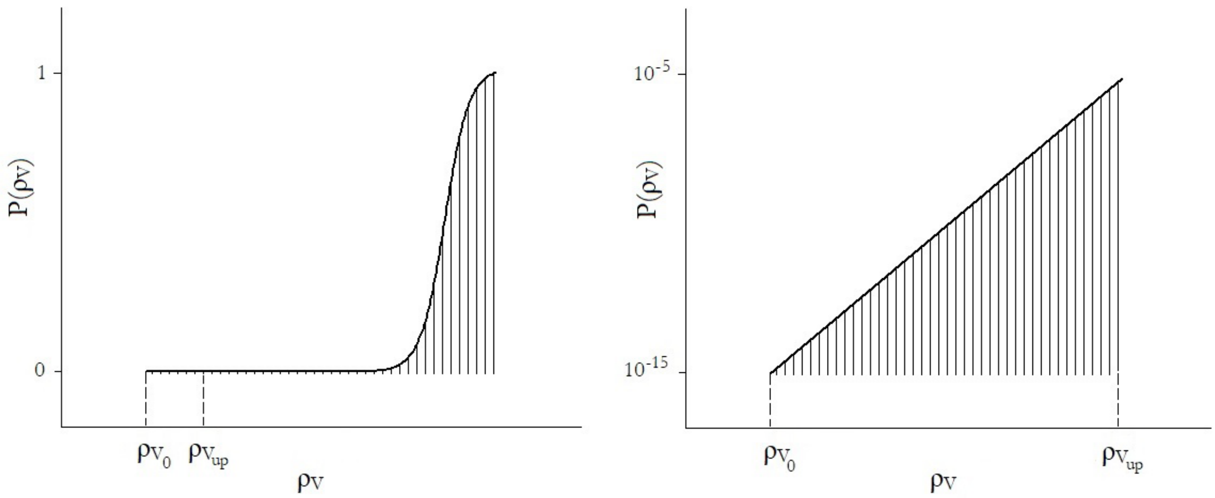

Figure 1

Hypothetical cumulative probability distribution of pv, showing an apparently constant range within observable limits (pv_0, pv_{up}) (left). The same observable range is zoomed in to show a variation of several orders of magnitude previously unnoticed due to scale comparisons (right)

Another key issue derived from the removal of P(*pv*) from eq. 1 is that it effectively became proportional to the number of observers in each universe. This introduced a subtle but crucial shift in how probabilities were assigned. Once observer abundance is given dominant weight in the posterior, the predictive distribution becomes primarily sensitive to multiplicity rather than to independently motivated variations in the prior occurrence probabilities. The methodological question, therefore, concerns how strongly posterior predictions should depend on observer abundance when prior structure remains weakly constrained.

This issue can be better understood by distinguishing between the two different components involved in anthropic probabilistic reasoning: the prior occurrence probabilities of possible universes and the way observer-related information is incorporated into predictive calculations. In a multiverse framework, one element of the prediction concerns how frequently universes with particular values of *pv* are expected to occur according to the underlying physical theory. A second element concerns how the presence and distribution of observers within those universes should influence the probability of what is ultimately observed.

Depending on how these two components are combined, predictive weight may be driven primarily by the prior likelihood of different universes or, alternatively, by the relative abundance of observers associated with each cosmological scenario. When observer abundance is given independent multiplicative weight, universes containing larger numbers of observers contribute more strongly to the posterior prediction, whereas approaches that rely primarily on prior occurrence probabilities place greater emphasis on the physical distribution of cosmological parameters themselves. The methodological question, therefore, is not whether observer information should be included —such conditioning is unavoidable in anthropic reasoning, as it is inherently *de se* in character— but how strongly predictive outcomes should depend on assumptions concerning the multiplicity of observers across possible universes.

In Weinberg's formulation, once the prior distribution P(*pv*) is assumed to be approximately uniform within the anthropically allowed range, it effectively cancels from the predictive expression. Then, the resulting probability distribution becomes proportional to the expected number of observers associated with each value of *pv*,

$$P_{obs}(pv) \propto N(pv)$$

Under these conditions, the posterior predictive distribution is driven primarily by observer abundance rather than on independently motivated differences in the prior occurrence probabilities of the corresponding universes. Structurally, this produces a weighting scheme closely related to those observer-count updating rules discussed in the literature on self-locating beliefs (Bostrom, 2003), in which scenarios with a greater number of observers receive a proportionally greater posterior weight. As a result, predictive outcomes become highly sensitive to assumptions concerning observer multiplicity when prior world-occurrence probabilities remain weakly constrained. From a cosmological perspective, this dependence raises the question of whether the apparent predictive success reflects genuine physical constraints on *pv* or, instead, actually expresses modelling assumptions about how observer populations are distributed across possible universes.

Weinberg's formulation implies that, under the assumption of a uniform prior distribution over possible values of *pv*, the posterior probability becomes proportional to the observer abundance. In this setting, we should expect to find ourselves in universes where observers are more frequent. However, this proportionality depends crucially on the simplification that prior occurrence probabilities do not differentially weight cosmological constant values. If the underlying distribution of universes were non-uniform, retaining the factor P(*pv*) in eq. 1 would play a substantive role in shaping the prediction. The appropriate treatment of this prior term remains a matter of ongoing debate, and arguments developed in related contexts (Bostrom, 2002b; Bradley, 2012) suggest that considering prior structure may strengthen the robustness of anthropic predictions concerning the cosmological constant.

To illustrate the role of observer-weighting more clearly, consider an adaptation of Leslie's emerald thought experiment (Leslie, 1996, p. 20). Suppose that, at an early stage in the experiment, three individuals are each given an emerald. Several centuries later, in a second stage, five thousand individuals are each given an emerald. Imagine that you find yourself holding an emerald but lack information about which stage of the experiment you belong to. If both stages are assumed equally likely a priori, self-locating reasoning suggests that you should assign a higher probability to belonging to the stage containing the larger number of participants, since there are more possible observer-locations compatible with your current evidence.

Since there are vastly more participants in the second century, a rational observer should conclude that it is overwhelmingly likely to belong to the later century. This probability is $P_{late} = 5000/5003 = 0.9994$, indicating that you are far more likely to be in the era of five thousand emerald recipients simply because it contains more observers.

Importantly, this inference does not imply that observers determine the underlying structure of the experiment. Rather, it reflects a probabilistic updating rule that conditions on indexical information —namely, the fact that one is an observer participating in the experiment. The resulting posterior probability depends both on the prior likelihood of each stage occurring and on the number of observer-locations associated with each stage.

This probabilistic reasoning reveals a structural parallel with Weinberg's prediction for *pv*. As eq. 1 shows, once P(*pv*) is treated as uniform, the posterior probability becomes proportional to the abundance of observers, so that self-location is effectively modelled as sampling from the largest population of observers. In this simplified setting, multiplicity alone drives the update. The analogy helps clarify this structural feature, but cosmological applications involve additional physical considerations. Unlike the *Emeralds* thought experiment, which abstracts from the conditions under which observers arise, cosmological modelling must also account for the underlying physical mechanisms that permit observer emergence. In this sense, the analogy is instructive but incomplete, since the predictive weight assigned to observer number ultimately depends on how those physical conditions are incorporated into the probabilistic framework.

To see this more clearly, suppose the thought experiment is modified: in addition to receiving an emerald, you also receive an undated astronomical report stating that there is a 50% chance that a catastrophic asteroid will strike Earth before that future distribution. If the asteroid hits, 90% of the future population will perish, and only five hundred survivors will remain to receive the emeralds. If you trust this report, then the probability of belonging to the second century must be revised downward to reflect that possibility. The probability of being in the second century should not be simply proportional to the number of observers but should also incorporate the probability of the catastrophic event occurring. So, the new calculation should be $P_{late} = (0.50\times5000+0.50\times500)/ (3+0.50\times5000+0.50\times500) = 0.9989$.

This simple example highlights the importance of relying on the frequency of world-observer pairs rather than uniquely on observer-relative frequencies[4]. By assuming uniform priors, Weinberg's approach becomes more dependent on observers multiplicity than on independently motivated variations in prior occurrence probabilities. Just as the asteroid impact affects the number of emerald holders in the second century, cosmological factors should also influence the probability of different universes existing —not just the number of observers within them. Thus, incorporating, and maintaining, this perspective should provide more robustness to the results.

2.2. Confronting H2: issues on conditionalization

A key aspect of Weinberg's anthropic argument is the conditionalization scheme he employed to predict the observed value of Λ. In this context, conditionalization refers to the process of weighting probabilities based on the presence of observers, ensuring that only habitable universes contribute to the final probability distribution of the measured Λ. Weinberg's assumption H2 stated that the number of observers, N, was proportional to the amount of baryonic matter that ended up in galaxies. This means that universes containing more galaxies should host more observers and, therefore, should be assigned a greater probability to obtain the prediction of *pv*.

This assumption appears *prima facie* reasonable: the formation of gravitationally bound structures, such as galaxies, is a necessary condition for life as we know it. However, a closer analysis reveals that this weighting scheme becomes problematic because no threshold for the formation of galaxies was well-defined. If any arbitrarily small number of galaxies could be considered sufficient for life, then Weinberg's conditionalization process may no longer constrain *pv* in a meaningful way.

The lack of a well-defined threshold for the formation of galaxies weakens the fine-tuning argument by allowing it to dissolve into a vague and imprecise selection effect, rather than a robust constraint on *pv*. In a stronger argument, the observed value of *pv* should be explained as falling within a narrowly constrained range that is necessary for the existence of observers. However, if the anthropic selection criterion is too flexible —allowing for an arbitrarily small number of galaxies to be considered viable for life— then the distinction between fine-tuned and non-fine-tuned universes becomes ambiguous. Instead of defining a clear boundary that separated universes where life could exist from those where it could not, the selection effect may have been applied too broadly, accommodating an indefinite range of *pv* values without offering a precise explanation for why we observe one particular value over another, other than being where the majority of observers are. Under such conditions, the argument risks losing the predictive framework that dictates the observed value of *pv*; instead, it becomes a *post hoc* rationalization that merely confirms that our universe must have a habitable *pv* because we exist to observe it.

To maintain the strength of the argument, it should be crucial to establish well-defined limits on the conditions necessary for observers to emerge. So, one of the major challenges in Weinberg's approach should have been to specify a minimum threshold for the formation of galaxies that would have meaningfully constrained *pv*. Consider the following scenario: as *pv* increases, the universe expands more rapidly, reducing the time available for matter to collapse into galaxies. But at some critical value of *pv*, only one galaxy forms; at a slightly lower value, two galaxies form; at an even lower value, three galaxies form, and so on. However, the number of observers does not necessarily scale linearly with these numbers of galaxies. If an undetermined number of galaxies are required before the first observer appears, then those universes on the higher end of the anthropic range may still be devoid of observers but still be factored into the calculations, potentially distorting the predicted distribution of *pv*.

---

[4] Or, equivalently, assuming uniform prior probabilities over world occurrence, such that the world-weighting factor cancels out in the probabilistic model and the posterior becomes effectively driven by observer abundance alone.

This introduces a fundamental inconsistency in Weinberg's conditionalization scheme: if some values of *pv* allowed the formation of galaxies but still resulted in uninhabited universes, then the assumption that more galaxies necessarily led to more observers becomes unreliable. In that case, the model would not account for scenarios where galaxies formed yet no observers emerged despite the presence of large-scale structures. This issue may not be purely theoretical, because when the predicted value using Weinberg's model was compared with the measured value of Λ, a discrepancy of three orders of magnitude was found (Bousso *et al.*, 2007). One possible explanation for this deviation may be the artificial inflation of the predicted peak of *pv* towards higher values, as explained above.

Building on the previous explanation, the anthropic range of *pv* could be better defined by adopting a more sophisticated model. A key enhancement could be the addition of thresholds for the emergence of life, as this process can be viewed as inherently probabilistic (observers ultimately arise from abiotic processes, and the transition from inert matter to organic molecules is highly stochastic in nature). An example that illustrates this point comes from RNA, a molecule thought to play a crucial role in the origin of life due to its ability to store information and self-replication. RNA is a chain of nucleotides that, given a specific sequence, can reproduce itself. Although the exact minimum length required for RNA to exhibit self-replicating properties remains uncertain, current estimates place it between 40 and 100 nucleotides (Totani, 2020). Based on these estimates, the probability of forming a self-replicating RNA sequence through random combinations can be calculated —and it is extraordinarily low. For instance, if we assumed a probability of such polymerization event of 1 in 10^{23} per billion years per star, then, considering that our universe contains roughly 10^{22} stars, the expected number of successful events that may eventually lead to the emergence of observers might be as low as one.

This calculation suggests that, in universes with fewer galaxies than our own, the expected number of observers may decrease far more sharply than a simple proportionality with baryonic collapse would imply. In Weinberg's framework, the number of observers is effectively treated as proportional to the fraction of matter that condenses into galaxies, so that

$$E[N(pv)] \propto F(pv)$$

where F(*pv*) denotes the fraction of matter forming gravitationally bound structures.

However, this assumption smooths over the stochastic structure of observer emergence. The transition from galaxy formation to the appearance of intelligent observers involves highly contingent biochemical and astrophysical processes that may introduce strong non-linearities into E[N(*pv*)]. If the probability of life emerging within a given galaxy is extremely small, then universes with modest galaxy formation could experience a disproportionately large decline in the expected number of observers.

Under such conditions, the relationship between cosmological parameters and observer abundance would not be well approximated by a linear scaling with galaxy mass. Instead, E[N(*pv*)] could exhibit threshold-like behaviour in expectation, even without imposing a discrete cutoff. In that case, the predictive distribution for *pv* may differ significantly from the one obtained under Weinberg's proportionality assumption.

The methodological concern, therefore, is not the need for an explicit threshold excluding universes with N=0, but whether the proxy used to estimate observer abundance captures the probabilistic structure of observer emergence with sufficient fidelity. If the emergence of observers is governed by rare and highly stochastic processes, then modelling E[N(*pv*)] as a smooth function of galaxy formation may artificially broaden the anthropic range and shift the predicted peak of *pv*.

This way, although the original model was functional to a certain extent, it should be reconsidered in favour of a more complex one. As it has been described (Bousso *et al.*, 2007), a star-based metric —expressed by the total stellar entropy— seems to be a more reliable predictor for the likelihood of observers than Weinberg's matter-forming galaxies. In addition, this approach may incorporate a more physically meaningful correlation between cosmic structure and the emergence of complex life.

2.3. Confronting H_4: issues on typicality

In cosmology, the typicality assumption plays a fundamental role in reasoning about observer-based probabilities. The Mediocrity Principle (MP), widely used in scientific literature, declares that our position in the universe should be considered representative rather than exceptional (Vilenkin, 1995). However, in a multiverse framework, typicality is not merely a philosophical stance, as it becomes a structural component of probabilistic inference.

Weinberg explicitly framed his anthropic reasoning in a way that preserved typicality once observer selection effects were taken into account. As he wrote:

> … the measured effective cosmological constant would be much smaller than the value expected on dimensional grounds in elementary particle physics, not because there is any physical principle that makes it small in all subuniverses, but because it is only in the subuniverses where it is sufficiently small that there would be anyone to measure it. (Weinberg, 1996, p. 3-4)

Immediately after, he reinforced the role of typical observers:

> In previous work I calculated the anthropic upper bound on the cosmological constant, which arises from the condition that *pv* should not be so large as to prevent the formation of gravitational condensations on which life could evolve. This bound is naturally larger than the average value of the cosmological constant that would be measured by typical observers, which obviously gives a better estimate of what we might find in our subuniverse. (Weinberg, 1996, p. 4)

Weinberg's prediction of Λ relied on the assumption that the number of observers scaled proportionally with the fraction of baryonic matter that condensed into galaxies (H2). Once this proportionality was adopted, an additional assumption was required in order to translate the abundance of observers into a probabilistic prediction. For this purpose, assumption H4 postulated that we are typical observers within a reference class that contains all such observers. In this way, the probability of observing a given value of *pv* becomes proportional to the total number of observers associated with that value.

This inferential step performs the precise function of specifying how self-locating uncertainty is resolved. Formally, it amounts to adopting a distribution over all observer-instances, such that

$$P(pv|\text{we observe}) \propto P(pv)N(pv) \qquad \text{eq. 2}$$

Typicality is therefore not merely a heuristic appeal to mediocrity, but a structural component of the probabilistic model. Without this assumption, the multiplicative weighting by N(*pv*) would not follow. If typicality were rejected or replaced by a non-uniform self-locating rule, the proportionality to N(*pv*) would not arise, and observer multiplicity would lose its direct predictive role in the posterior distribution.

However, the justification of the relevant reference class remains non-trivial. As Hartle and Srednicki (2007) emphasize, typicality should not be assumed by default but assigned only once a well-defined ensemble has been established. In cosmology, we lack empirical access to other intelligent observers, and thus the extension of the reference class beyond our own observational standpoint is necessarily theoretical. The MP licenses treating ourselves as typical within some ensemble, but it does not distinctly define how that ensemble ought to be characterized.

The essential question is not whether anthropic conditionalization is valid —it is inescapable— but rather how to address self-locating uncertainty within a *de se* framework. One common approach to resolving this issue involves considering ourselves as randomly selected from the entire collection of observer-instances, which means that universes with a higher number of observers are given proportionally more significance. Under an alternative resolution, the relevant self-locating fact is not which observer-instance we are among many, but simply that we inhabit a universe capable of supporting at least one observer. In that case, the predictive structure would instead resemble

$$P(pv \mid \text{we observe}) \propto P(pv)P(N(pv) \geq 1),$$

so that abundance beyond the minimal condition of observability would not automatically increase posterior probability. In the limiting case of extreme observer scarcity —where at most a single observer emerges in each viable universe— sampling over observer-instances would become substantially insignificant, since there would be no multiplicity over which probability weight could be distributed.

This reveals that Weinberg's conclusion depends on a specific resolution of self-locating uncertainty. The predictive force of N(*pv*) is conditional on adopting a uniform distribution over observer-instances. Alternative resolutions of *de se* uncertainty —such as conditioning only on the existence of observers rather than on our random selection among them— would not automatically generate proportional weighting by total observer number. Since anthropic reasoning itself does not uniquely determine how self-locating probabilities should be assigned, the choice requires independent justification.

These concerns connect with broader debates on self-locating beliefs. For example, weighting by the total number of observers, as in the Self-Indicating Assumption (SIA), tends to favour scenarios with larger populations and such reasoning may yield counterintuitive consequences when empirical constraints are weak (Bostrom, 2002a; Ćirković, 2004). The so-called Presumptuous Philosopher problem illustrates how probabilistic reasoning may grant overwhelming credence to theories that posit vastly larger numbers of observers, even in the absence of additional empirical support.

More generally, as Benétreau-Dupin (2015) observes in his analysis of cosmic probabilistic puzzles, the combination of neutrality assumptions —such as uniform priors— and appeals to typicality can yield surprisingly sharp predictions from limited informational input. The concern is that strong predictive conclusions may emerge from assumptions that were intended to express epistemic neutrality.

A further methodological refinement is suggested by Lacki (2021), who emphasizes that probabilistic outcomes associated with observer multiplicity depend crucially on how finely physical theories are decomposed into distinct micro-hypotheses. If observer-instances are treated as interchangeable without specifying the underlying physical microstates that differentiate them, the move from physical theory to probabilistic weighting remains underdetermined. In that case, multiplying by the total number of observers does not track a count of physically distinct realizations of the relevant observational situation, but instead reflects a coarse-grained representation in which qualitatively identical observer-instances are simply enumerated.

Similarly, Dorr & Arntzenius (2017) argue that the transition from qualitative priors over physical theories to self-locating beliefs requires substantive bridging principles that link physical hypotheses to indexical probability assignments. This reinforces the point that anthropic prediction depends not only on cosmological assumptions but also on how self-locating uncertainty is formally resolved.

In classical probability problems —such as drawing coloured balls from an urn— we reason as external agents with full knowledge of the sample space. The elements are discrete, empirically identifiable, and countable. Cosmology, by contrast, provides no external vantage point from which the total number and distribution of observers can be independently verified. While typicality arguments are highly effective in statistical mechanics, where macrostates are defined over well-characterized ensembles (McCoy, 2018), the cosmological application relies on theoretical postulates about observer abundance that remain empirically inaccessible.

Weinberg's framework is internally coherent once these assumptions are granted. Yet its predictive success depends not only on anthropic selection effects, but also on the adoption of a particular resolution of self-locating uncertainty and a particular level of theoretical coarse-graining. Since alternative, equally coherent resolutions remain available, the robustness of the prediction for *pv* should be understood as conditional rather than unconditional.

A central issue in Weinberg's framework concerns how self-locating uncertainty enters into the probabilistic reasoning. Once observer abundance is incorporated into the predictive expression, the resulting update depends on how indexical information should affect posterior probabilities. This structural feature finds an analogue in the Sleeping Beauty problem (Elga, 2000), which examines how an agent should revise her credence in a hypothesis upon receiving purely self-locating information. In that thought experiment, the disagreement does not concern the underlying physical setup, but the rule governing how multiplicity of observer-moments bears on rational belief revision. The debate is commonly divided into two principal positions:

— Thirders, who argue that the probability of heads should be 1/3, on the grounds that it leads to a single awakening (unlike tails, which results in two awakenings).

— Halfers, who maintain that the probability of heads remains 1/2, since the experiment does not change the initial chances, and the number of awakenings is irrelevant to the credence.

The philosophical significance of this debate extends beyond the thought experiment itself and bears on broader questions about self-locating probability in cosmology. The contrast between thirder and halfer positions illustrates how different resolutions of indexical uncertainty can yield different posterior weightings, even when the underlying physical hypotheses remain unchanged. In particular, thirders treat multiplicity of observer-moments as directly relevant to posterior probability, whereas halfers maintain that purely indexical information does not automatically warrant such proportional updating.

This structural contrast reflects the issue raised earlier. The role of multiple observers in probabilistic prediction depends on how observer instances are individualized and weighted. If observers cannot be meaningfully distinguished as distinct probabilistic units with respect to the data under consideration, their collective number may function as a modelling parameter, but its epistemic strength should require further justification.

From an epistemic standpoint, an additional concern arises regarding how ignorance about the total number of observers is represented. In the absence of empirical access to the distribution of observers across possible universes, our background information B appears compatible with a wide range of possibilities, that is, from a universe containing a single observer to one containing arbitrarily many. Formally, our state of knowledge in a background B does not privilege any specific disjunction of the form

$$N_1, N_1 \vee N_2, N_1 \vee N_2 \vee N_3, \ldots$$

where N_i denotes the hypothesis that exactly *i* observers exist. Mere ignorance about observer abundance does not, by itself, determine a probability distribution over these alternatives.

Once the Principle of Mediocrity is introduced, however, observer multiplicity becomes epistemically relevant. The probabilistic framework now includes weighting by observer count, thereby assigning greater posterior weight to hypotheses that contain larger numbers of observers. This shift does not follow from ignorance alone, but from the adoption of a substantive principle connecting observer abundance to self-location probability.

Norton (2010) has argued, in a related context, that certain forms of probabilistic reasoning can generate strong inferential shifts from evidentially neutral starting points. The present concern is not that Weinberg's reasoning is formally invalid, but that the transition from neutrality about observer number to multiplicity-sensitive updating requires an additional justificatory step. In this respect, anthropic predictions depend crucially on how ignorance about observer distribution is formalized within the probabilistic model. Since alternative resolutions of self-locating uncertainty remain available, the epistemic force of observer multiplicity should be understood as conditional on the adopted updating principle rather than as a direct consequence of empirical data alone.

3. Conclusions

This paper has presented a critical examination of the premises underlying Weinberg's probabilistic prediction of the observable cosmological constant. Within the framework of anthropic selection, Weinberg's reasoning is internally coherent and yielded an impressively accurate order-of-magnitude estimate when evaluated against the full theoretical range of possible values. His work remains a significant illustration of how selection effects may constrain otherwise weakly constrained cosmological parameters.

The analysis developed has highlighted that the predictive force of Weinberg's result depends not only on cosmological modelling, but also on how self-locating uncertainty is resolved within the probabilistic framework. Once observer abundance is incorporated into the predictive expression, additional epistemic commitments enter the model —specifically, principles governing how observer multiplicity bears on posterior probability assignments.

Therefore, the central issue is methodological. Certainly, anthropic conditionalization is unavoidable in any scenario where our existence functions as evidence. However, the transition from qualitative physical assumptions to quantitative beliefs about self-location requires transition principles that are not solely fixed by physical theory itself. In any case, different internally consistent resolutions of indexical uncertainty remain available, which can lead to varying degrees of multiplicity sensitivity in the resulting probability distribution.

This dependence indicates that the robustness of anthropic predictions depends largely on how the lack of knowledge about the observer distribution is formally represented and how observer instances are individualized within the model. Thus, the introduction of the number of observers as a weighting factor reflects substantial assumptions about self-location, rather than direct empirical constraints.

A more comprehensive treatment of these issues would therefore require further investigation into the principles linking cosmological hypotheses to indexical probability assignments, as well as clearer articulation of the level of theoretical granularity at which observer multiplicity is evaluated. Such refinements would clarify the epistemic status of the assumptions on which it depends.

In this sense, Weinberg's prediction should be understood not as a definitive confirmation of anthropic selection, but as a powerful illustration of how cosmological inference can hinge on subtle questions about self-locating probability. Recognizing the conditional structure of these assumptions allows for a more cautious yet conceptually transparent assessment of anthropic arguments in contemporary cosmology.

Acknowledgements

The author would like to express deep gratitude to Adan Sus and Maria Caamaño for their invaluable assistance in the preparation of this paper, and to an anonymous reviewer for their helpful and constructive comments.

REFERENCES

Azhar, F. (2014). Prediction and typicality in multiverse cosmology. *Classical and Quantum Gravity*, *31*(3), 035005. https://doi.org/10.1088/0264-9381/31/3/035005

Azhar, F. (2015). Testing typicality in multiverse cosmology. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, *91*(10), 103534. https://doi.org/10.1103/PhysRevD.91.103534

Azhar, F. (2016). Spectra of conditionalization and typicality in the multiverse. *Physical Review D*, *93*(4), 043506. https://doi.org/10.1103/PhysRevD.93.043506

Azhar, F. (2017). Three aspects of typicality in multiverse cosmology. In M. Massimi, J.-W. Romeijn, G. Schurz, & G. Weber (Eds.), *EPSA15 selected papers* (pp. 125-136). Springer. https://doi.org/10.1007/978-3-319-53730-6_11

Barrow, J. D., & Tipler, F. J. (1986). *The anthropic cosmological principle*. Oxford University Press.

Benétreau-Dupin, Y. (2015). Blurring out cosmic puzzles. *Philosophy of Science*, *82*(5), 879-891. https://doi.org/10.1086/683326

Bostrom, N. (2002a). *Anthropic bias: observation selection effects in science and philosophy*. Routledge.

Bostrom, N. (2002b). Self-locating belief in big worlds: Cosmology's missing link to observation. *Journal of Philosophy*, *99*(12), 607-623. https://doi.org/10.2307/3655771

Bostrom, N. (2003). The mysteries of self-locating belief and anthropic reasoning. *The Harvard Review of Philosophy*, *11*(1), 59-73. https://doi.org/10.5840/harvardreview20031114

Bousso, R., Harnik, R., Kribs, G. D., & Perez, G. (2007). Predicting the cosmological constant from the causal entropic principle. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, *76*(4), 043513. https://doi.org/10.1103/PhysRevD.76.043513

Bradley, D. (2012). Four problems about self-locating belief. *Philosophical Review*, *121*(2), 149-177. https://doi.org/10.1215/00318108-1539071

Carroll, S. M. (2004). Why is the universe accelerating? In R. E. Allen, D. V. Nanopoulos, & C. N. Pope (Eds.), *AIP conference proceedings* (Vol. 743, pp. 16-32). American Institute of Physics.

Carter, B. (1974). Large number coincidences and the anthropic principle in cosmology. *Symposium - International Astronomical Union*, *63*, 291-298. https://doi.org/10.1017/s0074180900235638

Ćirković, M. M. (2004). Is many likelier than few? A critical assessment of the self-indication assumption. *Epistemologia*, *27*(2), 265-298. https://doi.org/10.1400/23334

Elga, A. (2000). Self-locating belief and the Sleeping Beauty problem. *Analysis, 60*(266), 143-147. https://doi.org/10.1111/1467-8284.00215

Ellis, G. (2007). Issues in the philosophy of cosmology. In J. Earman & J. Buttefield (Eds.). *Handbook for the philosophy of physics* (pp. 1183-1285). Elsevier.

Freivogel, B. (2011). Making predictions in the multiverse. *Classical and Quantum Gravity*, *28*(20), 204007. https://doi.org/10.1088/0264-9381/28/20/204007

Garriga, J., & Vilenkin, A. (2000). On likely values of the cosmological constant. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, *61*(8), 083502. https://doi.org/10.1103/PhysRevD.61.083502

Hartle, J. B., & Srednicki, M. (2007). Are we typical? *Physical Review D - Particles, Fields, Gravitation and Cosmology*, *75*(12), 123523. https://doi.org/10.1103/PhysRevD.75.123523

Lacki, B. C. (2023). The noonday argument: fine-graining, indexicals, and the nature of copernican reasoning. *International Journal of Astrobiology*, *22*(4), 354-398. https://doi:10.1017/S1473550423000071

Leslie, J. (1996). *The end of the world: the science and ethics of human extinction*. Routledge.

Martel, H., Shapiro, P. R., & Weinberg, S. (1998). Likely values of the cosmological constant. *The Astrophysical Journal*, *492*, 29-40.

McCoy, C. D. (2018). The implementation, interpretation, and justification of likelihoods in cosmology. *Studies in History and Philosophy of Science Part B - Studies in History and Philosophy of Modern Physics*, *62*, 19-35. https://doi.org/10.1016/j.shpsb.2017.05.002

Norton, J. D. (2010). Cosmic confusions: not supporting versus supporting not. *Philosophy of Science*, *77*(4), 501-523. https://doi.org/10.1086/661504

Page, D. N. (2008). Cosmological measures without volume weighting. *Journal of Cosmology and Astroparticle Physics*, *2008*(10), 025. https://doi.org/10.1088/1475-7516/2008/10/025

Polchinski, J. (2005). The cosmological constant and the string landscape. In D. Gross, M. Henneaux, & A. Sevrin (Eds.), *The quantum structure of space and time. Proceedings of the 23rd Solvay conference on physics* (pp. 215-233). World Scientific Publishing Co. Re. Ltd.

Tegmark, M., Aguirre, A., Rees, M. J., & Wilczek, F. (2006). Dimensionless constants, cosmology, and other dark matters. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, *73*(2), 023505. https://doi.org/10.1103/PhysRevD.73.023505

Totani, T. (2020). Emergence of life in an inflationary universe. *Scientific Reports*, *10*(1), 1671. https://doi.org/10.1038/s41598-020-58060-0

Vilenkin, A. (1995). Predictions from quantum cosmology. *Physical Review Letters*, *74*(6), 846-849. https://doi.org/10.1103/PhysRevLett.74.846

Vilenkin, A., & Yamada, M. (2020). Four-volume cutoff measure of the multiverse. *Physical Review D*, *101*(4), 043520. https://doi.org/10.1103/PhysRevD.101.043520

Weinberg, S. (1987). Anthropic bound on the cosmological constant. *Physical Review Letters*, *59*(22), 2607-2610. https://doi.org/10.1103/PhysRevLett.59.2607

Weinberg, S. (1989). The cosmological constant problem. *Reviews of Modern Physics*, *61*(1), 1-23. https://doi.org/10.1142/9789812704030_0315

Weinberg, S. (1996). Theories of the cosmological constant. In N. Turok (Ed.), *Conference on critical dialogs in cosmology* (pp. 195-203). World Scientific.

Weinberg, S. (2000). A priori probability distribution of the cosmological constant. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, *61*(10), 103505. https://doi.org/10.1103/PhysRevD.61.103505

Eduard Dàvila is a researcher in philosophy of cosmology affiliated with the University of Valladolid. His work focuses on anthropic reasoning, typicality assumptions, and probabilistic reasoning in multiverse cosmology.

Address: Department of Philosophy, University of Valladolid, Plaza Campus Universitario, s/n, 47011 Valladolid, Spain. E-mail: eduard.davila@estudiantes.uva.es – ORCID: https://orcid.org/0000-0001-5419-8591

Theoria, 2026, 41(1), 87-102
https://doi.org/10.1387/theoria.27281

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

¿Hay un sesgo de financiación en la investigación biomédica?

(Is there a funding bias in biomedical research?)

Francesc Roig-Loscertales*
UNED
https://orcid.org/0000-0002-3585-7536

Palabras clave

Sesgo
Sesgo de financiación
Industria farmacéutica
Colaboración Cochrane
Investigación clínica

RESUMEN: El hallazgo de que los estudios financiados por la industria farmacéutica suelen reportar con mayor frecuencia resultados favorables al promotor ha llevado a proponer la existencia de un sesgo de financiación que podría sobreestimar los efectos de los medicamentos. Sin embargo, no se ha identificado un mecanismo claro ni hay evidencia sólida de que la financiación de la industria introduzca en los resultados un error sistemático favorable a sus intereses. Este artículo revisa críticamente el concepto de sesgo de financiación, argumentando que la influencia de la industria no debe considerarse un sesgo específico. Para ello, a partir de la definición de sesgo adoptado por la metodología Cochrane – el *patrón oro* en evaluación de intervenciones en salud- analiza qué características debería cumplir la influencia de la industria para poder mantener que produce sesgo en la investigación clínica y, a través de una caso empírico, pone de manifiesto cómo no puede asumirse acríticamente que los resultados de los estudios que financia y en los que tiene obvios intereses comerciales obtienen resultados diferentes a los financiados por otras fuentes. Por ello, mantiene como conclusión la necesidad de abandonar la idea de sesgo de financiación, evitando la estigmatización de la investigación privada por su origen, como si de un pecado original se tratara.

Keywords

Bias
Funding bias
Pharmaceutical industry
Cochrane Collaboration
Clinical research

ABSTRACT: The observation that studies funded by the pharmaceutical industry tend to report results more frequently favorable to the sponsor has led to the suggestion of a funding bias that could overestimate the effects of medications. However, no clear mechanism has been identified, and there is a lack of compelling evidence that industry funding introduces a systematic error in the results that favors its interests. This article provides a critical review of the concept of funding bias, arguing that industry influence should not be regarded as a distinct form of bias. Drawing on the definition of bias established by the Cochrane methodology —widely considered the *gold standard* in the evaluation of health interventions— the analysis explores the characteristics that industry influence would need to meet in order to be considered a source of bias in clinical research. Using an empirical case study, it is demonstrated that one cannot uncritically assume that studies funded by the industry and aligned with its commercial interests produce different results than those funded by other sources. Consequently, the article asserts the need to abandon the notion of funding bias and to avoid stigmatizing private research based solely on its origin, as if it were a form of original sin.

* ***Correspondence to:*** Francesc Roig-Loscertales. Dpto. de Lógica, Historia y Filosofía de la ciencia, UNED. Paseo de Senda del Rey 7, 28040 Madrid, España – froig9@alumno.uned.es – https://orcid.org/0000-0002-3585-7536

How to cite: Roig-Loscertales, Francesc (2026). «¿Hay un sesgo de financiación en la investigación biomédica?»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 41(1), 87-102. (https://doi.org/10.1387/theoria.27281).

Received: 12 February, 2025; Final version: 28 November, 2025.
ISSN 0495-4548 - eISSN 2171-679X / © 2026 UPV/EHU Press

Gako-hitzak

Alborapena
Finantzaketa-alborapena
Farmazia-industria
Cochrane Lankidetza
Ikerketa klinikoa

LABURPENA: Farmazia-industriak finantzatutako azterketek babeslearen aldeko emaitzak eman ohi dituztela ikusita, iradoki da finantzaketa-alborapen bat dagoela eta horrek sendagaien efektuak neurriz kanpo balioestea ekar dezakeela. Hala eta guztiz, ez da mekanismo argirik identifikatu eta ez dago ebidentzia sendorik esateko ezen industriaren finantzaketak beraren interesen aldeko errore sistematiko bat gehitzen duenik emaitzetan. Artikulu honek *finantzaketa-alborapen* kontzeptua berrikusten du kritikoki, eta argudiatzen du industriaren eragina ez dela aparteko alborapen-modutzat hartu behar. Cochrane metodologiak ezarritako alborapenaren definizioa abiapuntutzat hartuta —askok *urre-patroitzat* jotzen dute osasun-arloko esku-hartzeen ebaluazioan—, artikuluak aztertzen du zer ezaugarri izan beharko lituzkeen industriaren eraginak ikerketa klinikoaren arloko alborapen-iturritzat jo dadin. Kasu enpiriko baten azterketatik abiatuta, artikuluak agerian uzten du ezin dela akritikoki ziurtzat jo industriak finantzatutako eta industriaren interes komertzialekin lerrokatutako ikerketek emaitza desberdinak ematen dituztela beste iturri batzuek finantzatutako ikerketekin alderatuta. Beraz, artikuluak aldezten du finantzaketa-alborapenaren ideia alde batera utzi behar dela, eta saihestu egin behar dela ikerketa pribatua bere jatorriagatik estigmatizatzea, jatorrizko bekatu moduko bat balitz bezala.

1. Introducción

En una revisión sistemática Cochrane de 2012, Lundh y colaboradores hallaron que la investigación clínica financiada por la industria farmacéutica estaba asociada a una mayor probabilidad de resultados favorables a sus productos, en comparación con los estudios que tenían otras fuentes de financiación (Lundh *et al.*, 2012). El metanálisis confirmaba lo que diferentes revisiones previas ya señalaban, esta vez con la robustez de la metodología Cochrane. Cinco años después, una actualización de los mismos autores corroboró el resultado (Lundh *et al.*, 2017) lo que confirmó su propuesta de considerar la financiación por la industria como un nuevo tipo de sesgo específico*, el sesgo de financiación* (Lundh *et al.*, 2017, p. 20). En realidad, para una parte de la comunidad médica el resultado no representaba ninguna sorpresa, simplemente aportaba la evidencia definitiva sobre un hecho que daba por supuesto: la industria farmacéutica influye sobre la investigación clínica para obtener resultados de acuerdo a sus intereses comerciales. El efecto que provocaría la financiación de la industria sobre la generación de conocimiento científico ha sido calificado por numerosos filósofos y sociólogos de la ciencia como *sesgo de financiación*[1] (entre otros, Holman & Elliot, 2018; Reutlinger, 2020; Solomon, 2020; Leefmann, 2021; Sismondo, 2021).

El trabajo que se presenta revisa críticamente la existencia de este presunto *sesgo de financiación* y mantiene que la influencia de la industria en la investigación clínica no puede ser considerada como un sesgo específico conforme a los criterios usados en investigación clínica. Para ello, en la sección siguiente se examina el llamado de *sesgo de financiación*, revisando la evidencia existente sobre la influencia de la industria en la investigación biomédica, los supuestos mecanismos a través de los cuales se ejercería y las acciones propuestas para contrarrestarlo. En la tercera sección se presenta la metodología de la Colaboración Cochrane, considerada de facto el *patrón oro* en la evaluación de la evidencia biomédica. Contra la recepción entusiasta de este supuesto sesgo de financiación en la filosofía y sociología de la investigación biomédica, en el cuarto apartado se defiende que el concepto de *sesgo*, tal y como se entiende en investigación clínica, no puede utilizarse para caracterizar los efectos de esta influencia. Finalmente, en las conclusiones se propone abandonar el concepto de sesgo de financiación como algo diferenciado y específico de la investigación patrocinada por la industria farmacéutica y profundizar en cómo actúa caso por caso la influencia de la financiación privada.

1 El sesgo de financiación recibe en la literatura en inglés diferentes nombres: *funding bias*, *sponsorship bias* e *industry bias.*

2. El sesgo de financiación

Ya a principios del siglo XX, Torald Sollmann señalaba el problema que podía originar la cada vez mayor presencia de la industria farmacéutica en la investigación médica. Miembro del *Council on Pharmacy and Chemistry of the American Medical Association* (antecedente de la FDA), Sollmann advertía de los efectos que podía conllevar la opacidad y el secreto en el desarrollo y evaluación de fármacos y la publicación de resultados, y señalaba la poca calidad y el secretismo como las mayores amenazas a la credibilidad de los informes enviados al *Council* (Sollmann, 1917). Sin embargo, la investigación empírica sobre los efectos provocados por el patrocinio de la investigación por la industria farmacéutica no empezó hasta tiempo después, cuando Elina Hemminki (1980) analizó por primera vez el sesgo de información (*reporting bias*) en los documentos regulatorios, evaluándolos comparativamente con el contenido de los informes. A partir del estudio de 566 solicitudes de autorización de fármacos psicotrópicos en Finlandia y Suecia en los años 1965, 1970, 1974 y 1975, la autora observó que la mayoría de los ensayos clínicos que incluían no estaban controlados y eran de mala calidad, que muchos no habían sido publicados previamente por lo que su diseño y resultados no estaban a disposición de la comunidad médica, y que la información que contenían era, a menudo, deficiente. El origen del problema se hallaría en aquellos estudios que se llevan a cabo por motivos diferentes a los del avance del conocimiento (por ejemplo, para obtener licencia de comercialización, ganancias económicas, en beneficio de la propia carrera profesional, etc.), hecho que provocaría que tuvieran una mayor probabilidad de aparición de sesgos (Jefferson, 2020).

Hay considerable evidencia sobre la existencia de asociación entre financiación de la industria farmacéutica y notificación de resultados favorables al promotor en diferentes tipos de estudios clínicos (por ejemplo, y entre otros, Bekelman *et al.*, 2003; Lexchin *et al.*, 2003; Schott *et al.*, 2010; Flacco *et al.*, 2015). Esta asociación, y de forma general el hecho de que la financiación de los estudios por la industria pueda influir en los resultados de investigación, ha generado preocupación en la comunidad médica (Gazendam *et al.*, 2022), incluso ha sido calificada como la mayor amenaza a la objetividad de la investigación biomédica (Solomon, 2020). Sin duda, la evidencia más robusta de la asociación entre financiación de la industria y resultados es la proporcionada por Lundh y colaboradores (2012; 2017) en una revisión sistemática siguiendo la metodología Cochrane. El objetivo de la revisión era doble: en primer lugar, evaluar si existía asociación entre el patrocinio de la industria farmacéutica y resultados favorables al patrocinador; en segundo lugar, si los estudios sobre fármacos y dispositivos patrocinados por la industria tenían un riesgo de sesgo diferente que los estudios que tenían otras fuentes de patrocinio. Para ello, se seleccionaron revisiones publicadas que comparaban estudios patrocinados por la industria con estudios que tenían otras fuentes de financiación y revisiones que comparaban estudios de productos de fabricantes competidores. Sus resultados son ampliamente conocidos: los estudios patrocinados por la industria conseguían resultados favorables con mayor frecuencia que los estudios no patrocinados por la industria[2]. En cambio, no se encontraron diferencias en el riesgo de sesgo según patrocinador, excepto en el proceso de cegado, donde los estudios de la industria presentaban menor riesgo. A partir de los resultados, los autores sugerían la existencia de un sesgo específico provocado por la financiación de la industria farmacéutica que no podría explicarse mediante las evaluaciones estándar. Dado que los resultados más favorables estarían mediados por factores distintos de los documentados en la metodología Cochrane nos hallaríamos ante un tipo nuevo de sesgo, con sus mecanismos particulares. Como la metodología de los metanálisis Cochrane se considera la más robusta para revisiones de la evidencia, estos resultados se han considerado a menudo como confirmación definitiva de la existencia de un *sesgo de financiación* (*funding bias* o *sponsorship bias*). De hecho, los autores proponían en las conclusiones que el patrocinio de la industria debería conside-

---

2 En los resultados los autores señalan que en comparación con los estudios no patrocinados por la industria, los estudios patrocinados tuvieron con mayor frecuencia resultados favorables de eficacia [con un cociente de riesgo (RR) de 1,27 (IC del 95%: 1,17 a 1,37) (25 artículos) (evidencia de calidad moderada)]; resultados similares de efectos secundarios, [RR: 1,37 (IC del 95%: 0,64 a 2,93) (cuatro artículos) (evidencia de muy baja calidad)] y presentaron más frecuentemente conclusiones favorables, [RR: 1,34 (IC del 95%: 1,19 a 1,51) (29 artículos) (evidencia de baja calidad)] (Lundh, 2017).

rarse como inductor de sesgo específico, el sesgo de financiación (Lundh *et al.*, 2017, p. 20). Otros autores también proponen que se incorpore en el *Cochrane Risk of Bias Tool* (Bero, 2013).

Una vez probada la existencia de sesgo, los esfuerzos se han dirigido a la identificación y caracterización de los mecanismos a través de los cuales alterarían los resultados. Hoy por hoy, son dos los enfoques principales para identificarlos. En primer lugar, están los que adoptan una concepción atomista[3], aquellos para quienes el sesgo de financiación se produciría a través de mecanismos particulares que aparecerán puntualmente en algunos estudios. Por ejemplo, a través de la elección de comparadores y dosis (Rochon *et al.*, 1994; Safer, 2002; Sinyor *et al.*, 2012; Gøtzsche, 2013), la forma de administración de los fármacos (Johansen & Gøtzsche, 1999), la elección de variables (Djulbegovic *et al.*, 2013), la codificación y análisis de los datos (Juni & Dieppe, 2002; Gøtzsche, 2013) o la notificación selectiva de resultados y el sesgo de publicación (Vedula *et al.*, 2009). Es decir, los intereses del patrocinador, como causa próxima, podrían llevar en algunas ocasiones, a través de distintos mecanismos, a la desviación sistemática del resultado de los ensayos. Una segunda concepción, que podríamos calificar de *holista*, es la que ha defendido el sociólogo Sergio Sismondo (2018, 2021), uno de los autores de la revisión Cochrane que propone introducir el sesgo de financiación como un tipo de sesgo específico (Lundh *et al.*, 2017, p. 20). La perspectiva holista mantiene que los efectos de la financiación de la industria se deberían a la suma de múltiples acciones que tendrían como resultado una influencia indeterminada, siempre favorable a los intereses de la industria pero difícil de identificar y cuantificar. En el caso de Sismondo, su tesis central es que la industria utiliza redes de gestión de la investigación, publicación y difusión de conocimiento que estarían ocultas a la comunidad científica, así como a reguladores y decisores, consiguiendo que el conocimiento resultante tenga la apariencia de «neutral» (Sismondo, 2018, p. 178). Las empresas subcontratadas para realizar los ensayos (CROs, en su acrónimo inglés) y realizar su análisis y difusión organizarán el proceso en función de los intereses de su cliente, de modo indetectable para los análisis atomistas de los sesgos, pues la manipulación se realizaría cumpliendo con sus directrices metodológicas (aleatorizar, cegar, etc.). Fernández-Pinto (2023) mantiene por su parte que la influencia de la industria se ejercería a través de lo que califica como *sesgos implícitos*: a través de múltiples canales el patrocinador crearía una predisposición inconsciente en los investigadores a tomar las decisiones que más le convengan a la industria a lo largo del proceso de investigación.

Se han propuesto diferentes estrategias para evitar o corregir los sesgos que la financiación privada produce sobre los resultados de investigación. Un primer grupo de medidas pivota sobre la transparencia, y se fueron generalizando en los 2000. En primer lugar, exigir a los autores de un manuscrito que revelen sus potenciales conflictos de intereses, enumerando los patrocinadores del estudio para alertar a evaluadores y lectores sobre sus posibles sesgos. En segundo lugar, la obligación de incluir cualquier ensayo clínico que se vaya a realizar en un registro públicamente accesible, gestionado por los reguladores, al que se deben añadir posteriormente los resultados obtenidos, para evitar así manipulaciones *ad hoc* (Laine *et al.*, 2007; Tsé & Zarin, 2009). Por último, la estandarización del propio proceso: las guías CONSORT para la comunicación de ensayos clínicos (publicada originalmente en 1996 (Begg *et al.*, 1996) con diversas actualizaciones posteriores) y STROBE para los estudios observacionales (publicada originalmente en 2007 (Vandenbroucke *et al.*, 2007) pretenden unificar la información que debe comunicarse en la publicación de resultados, facilitando así la identificación de posibles sesgos por parte de evaluadores y lectores. No obstante, y a pesar de que la investigación patrocinada por la industria suele cumplir con todas estas políticas igual o mejor que la investigación pagada con fondos públicos (Goldacre *et al.*, 2018, en Europa; De Vito *et al.*, 2020 y Piller, 2020 en Estados Unidos), la asociación entre financiación privada y resultados sigue presente. Esto ha sido interpretado de distinto modo según la concepción del sesgo de financiación. Para los atomistas, es un recordatorio de que no toda investigación con patrocinio privado estará sesgada (Resnik & Elliot, 2013). Para los partidarios de la concepción holista, esto probaría la impotencia de todas estas políticas para corregir el sesgo, proponiendo distintas alternativas que intervienen, en orden creciente, sobre la investigación privada.

[3] En el articulo se utilizan los términos «atomista» y «holista» exclusivamente para diferenciar dos estrategias que intentan explicar un mismo fenómeno: concretamente, la influencia de la industria farmacéutica en los resultados de investigación de los estudios clínicos que patrocina. Queda más allá del objetivo de este trabajo indagar sobre actividades como la influencia sobre reguladores, etc., que, aunque sin duda muy interesantes, requieren de otro tipo de aproximación. Agradecemos a uno de los revisores la necesidad de esta aclaración.

Así, hay quien sugiere intervenir sobre los resultados y corregirlos para eliminar el sesgo. Miriam Solomon (2020) ha propuesto aplicar una corrección cualitativa y una cuantitativa a todos los estudios financiados por la industria. La primera supondría diferenciar dos escalones en la cima de la pirámide de la evidencia: en el primero estarían los Ensayos Clínicos Aleatorizados (ECA, de ahora en adelante) con patrocinio público, en el segundo los ECA con patrocinio privado. La segunda se aplicaría solo al resultado de estos últimos y supondría un factor de corrección, ponderando el efecto del tratamiento por la sobre-estimación que se observa en promedio en los ECA patrocinados.[4] También hay quien propone intervenir sobre los propios procesos de investigación. Fernández-Pinto (2023) sugiere aplicar técnicas de corrección de los sesgos implícitos, entrenando a los investigadores para detectarlos y eliminarlos. Finalmente, están quienes desconfían de reformas parciales como las dos propuestas anteriores, y proponen eliminar completamente la financiación de la industria en la investigación clínica y que se lleve a cabo únicamente a través de organizaciones independientes, normalmente gubernamentales, que puedan financiarse mediante un impuesto a las industrias interesadas en el desarrollo de fármacos (Krimsky, 2003; Angell, 2005; Sismondo, 2008; Gotzsche, 2013).

Para poder evaluar estas propuestas, debemos reconsiderar la justificación de la concepción atomista de los sesgos que sus críticos holistas pretenden superar. Para ello, examinaremos ahora la formulación canónica de los sesgos en investigación clínica que debemos a la Colaboración Cochrane. Después, discutiremos si de acuerdo con el patrón oro en investigación clínica es posible hablar de un sesgo de financiación utilizando un caso real.

3. *El enfoque Cochrane sobre los sesgos en la investigación clínica*

Desde hace ya tres décadas, la Colaboración Cochrane pone a disposición del público síntesis de la mejor evidencia disponible sobre tratamientos médicos, mediante revisiones sistemáticas de los ensayos clínicos aleatorizados (ECA) que evalúan su seguridad y eficacia (Cumpston *et al.*, 2024). El objetivo de la Cochrane es producir información médica fiable, libre de conflictos de intereses, sobre la que sea posible desarrollar guías de práctica clínica y políticas sanitarias (Bero *et al.*, 2018). La Colaboración Cochrane se ha convertido así en la principal fuente internacional de *revisiones sistemáticas* (Flemyng *et al.*, 2023), gracias a una metodología rigurosa y estandarizada descrita en el *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins *et al.*, 2024). La fiabilidad de una revisión sistemática depende principalmente del control sobre los sesgos que pudieran contener los ensayos. Los sesgos, para la Cochrane, son interferencias en el experimento que producen diferencias sistemáticas entre sus resultados y los verdaderos efectos del tratamiento (Lundh & Gotzsche, 2008). Por este motivo, la Cochrane ha desarrollado una herramienta específica para evaluar el riesgo de sesgo en los ECA denominada *Cochrane Risk of Bias* (de ahora en adelante, *RoB*), a través de la cual se ha sistematizado el propio concepto de sesgo en la investigación biomédica.

En la *RoB* de la Cochrane el sesgo se define como «un error sistemático, o desviación de la verdad, en los resultados» (Boutron *et al.*, 2024)[5]. Un error sistemático es aquel que provocaría que múltiples repeticiones de un mismo estudio dieran, por término medio, una respuesta equivocada. Para que una característica determinada pueda ser considerada como generadora de sesgo deben cumplirse dos condiciones: en primer lugar, debe identificarse y describirse el mecanismo a través del cual origina esa «desviación sistemática»; en segundo lugar, debe probarse empíricamente su efecto. Solo así es posible asegurar que el error no puede tener otro tipo de explicaciones (Boutron *et al.*, 2024). Así, un pro-

---

[4] Solomon (2020, p. 448) describe la aplicación práctica de esta corrección cuantitativa a partir de un ejemplo con un hipotético ECA sobre un nuevo fármaco patrocinado por la industria que obtiene un resultado según el cual disminuye el dolor de forma 3 veces más probable que el placebo. Suponiendo que la medida de OR más reciente para el sesgo de la industria es la de Lundh (2017) de 2,05. Entonces, los resultados ajustados para el fármaco de la industria serían 3 x 1/2,05 = 1,46.

[5] Omitiremos la discusión de conceptos alternativos de sesgo en la investigación biomédica de uso común en filosofía, pues estos conceptos no han tenido apenas influencia en la definición y debate del sesgo de financiación –aunque algunos de ellos reaparecerán más adelante en la discusión.

https://doi.org/10.1387/theoria.27281

ceso bien conocido donde existe riesgo de sesgo es en la selección y asignación de los pacientes a los grupos del estudio. La asignación puede realizarse de forma aleatoria, donde es el azar el que determina el grupo de tratamiento, o bien de forma no-aleatoria, donde es el investigador quien asigna según su criterio el paciente a un grupo u otro. Es sabido que en el modelo no-aleatorio existe el riesgo de que se produzca el llamado *sesgo de* selección provocando una distorsión de los resultados: en lugar de medir el efecto del tratamiento sobre el paciente medio, se mediría sobre el tipo de paciente seleccionado por el investigador conforme a sus preferencias. El mecanismo que generaría este error sistemático es la decisión del investigador, que suele efectuar el reclutamiento y asignación de los pacientes en función de la presencia o ausencia de determinadas características. Realizada adecuadamente, la aleatorización impide que las preferencias del investigador afecten al proceso de medida y permiten, mediante el análisis de la varianza, estimar sus efectos (Higgins *et al.*, 2024). No obstante, como se ha señalado, no basta con identificar y describir el mecanismo por el que se origina el sesgo: es preciso probar empíricamente que su presencia provoca un error sistemático en los resultados. En el caso de la asignación de los pacientes, hay evidencia sólida de que los resultados de los estudios no aleatorizados pueden sobrestimar de forma espuria los beneficios del tratamiento, dando lugar a conclusiones engañosas. Así ocurrió por ejemplo en el caso de los efectos de la terapia anticoagulante en el infarto agudo de miocardio (Chalmers *et al.*, 1977), el de la terapia hormonal sustitutiva y el riesgo de enfermedad coronaria (Hulley *et al.*, 1998), el del impacto en la mortalidad coronaria del betacaroteno y el alfa tocoferol (The Alpha Tocopherol Beta Carotene Cancer Prevention Study Group, 1994; Yusuf *et al.*, 2000) o en la relación entre la fibra dietética y el cáncer de colon (Schatzkin *et al.*, 2000). En todos ellos, los resultados obtenidos fueron superiores cuando los pacientes se asignaron de forma no-aleatoria.

Además de probar empíricamente que se alteran los resultados, ante una posible sospecha de elemento generador de sesgo la Cochrane requiere también que se descarten posibles explicaciones alternativas. Ejemplos bien conocidos son la asociación entre el tamaño de la muestra y las estimaciones del efecto del tratamiento, cuando se observan efectos mayores en ensayos con muestras más pequeñas (Dechartres *et al.*, 2013), y la asociación entre ensayos realizados en un solo centro y efectos mayores que en los ensayos multicéntricos, incluso después de controlar el tamaño de la muestra (Dechartres *et al.*, 2011). En ambos casos la asociación podría deberse al sesgo de publicación, ya que los estudios con mayores efectos quizá tengan más probabilidades de publicarse. Sin embargo, pueden existir explicaciones alternativas: los participantes en ensayos más pequeños o en un solo centro pueden ser más homogéneos, generando efectos mayores. Por tanto, para la Cochrane en estos casos no se puede hablar de sesgo con certeza (Boutron *et al.*, 2024). Este es un aspecto esencial de la herramienta *RoB*: no es automática, el autor de cada revisión debe emitir un juicio sobre si es probable que la forma en que se ha llevado a cabo el ensayo ha provocado riesgo de sesgo. Por ello, la *RoB* establece que el fundamento de las evaluaciones de riesgo de sesgo debe hacerse explícito, registrando junto al juicio los aspectos metodológicos en los que se basa. De esta forma se proporciona una mayor transparencia, permitiendo a los lectores decidir si están de acuerdo con los juicios realizados (Higgins *et al.*, 2011).

La *RoB* está organizada en categorías llamados *dominios* que cubren todos los tipos de sesgo que actualmente se considera pueden afectar a los resultados de los ECAs. Estos dominios son: sesgo derivado del proceso de aleatorización; sesgo debido a desviaciones de las intervenciones previstas; sesgo debido a la falta de datos sobre el resultado; sesgo en la medición del resultado; y sesgo en la selección del resultado comunicado, reservándose un sexto dominio para «otros sesgos». En la evaluación del riesgo de sesgo es obligatorio que el autor considere todos los dominios, y que no añada ningún otro. Para cada dominio, la herramienta comprende una serie de *preguntas de señalización* que proporcionan un enfoque estructurado que permite obtener la información relevante para la evaluación. Las preguntas pretenden ser razonablemente objetivas, aunque la misma herramienta admite que algunas pueden requerir cierto grado de juicio. Además, cada dominio comprende también un juicio sobre el riesgo de sesgo para el ámbito, facilitado por un algoritmo que relaciona las respuestas a las preguntas de señalización con un juicio propuesto; cuadros de texto libre para justificar las respuestas a las preguntas de señalización y los juicios sobre el riesgo de sesgo; y una opción para predecir (y explicar) la dirección probable del sesgo (Sterne *et al.*, 2019). En la aplicación práctica de la herramienta, cada estudio debe ser evaluado de forma independiente por lo menos por dos revisores diferentes, que luego deben ponerse de acuerdo en el juicio, y el proceso de evaluación del riesgo de sesgo debe ser transparente, de forma que los lectores deben ser capaces de discernir por qué un resultado concreto se calificó de bajo riesgo de sesgo y por qué otro se calificó de alto riesgo de sesgo (Boutron *et al.*, 2024). Además del riesgo de sesgo de cada estudio, los autores de las revisiones deben valorar cómo afectan los sesgos de los estudios a los resultados del metanálisis, de forma que sea posible determinar la solidez de las conclusiones.

Cualquier revisión sistemática de la Cochrane permite visualizar cómo se lleva a cabo el análisis del riesgo de sesgo. Es posible tomar como ejemplo la que evaluó el efecto de las estatinas en la prevención primaria de la enfermedad cardiovascular (Taylor *et al.*, 2013). Una vez definidos los criterios de inclusión para los estudios y seleccionados aquellos que los cumplían, los autores realizaron la evaluación del riesgo de sesgo de cada uno a través de la *RoB*. Así, para el ECA que evaluaba los efectos de la atorvastatina (Bone *et al.*, 2007), el análisis de riesgo de sesgo en el dominio «generación aleatoria de la secuencia» fue a juicio de los autores «bajo», justificándolo en que se había usado un «código pseudoaleatorio generado por ordenador». En el «reporte selectivo» (dominio) también fue «bajo» (juicio de los autores) porque «todos los resultados se habían comunicado» (justificación del juicio). En cambio, en el dominio del «cegamiento» el riesgo era «poco claro» (juicio de los autores) porque «si bien se afirma que es doble ciego solo se informa de que los pacientes desconocen el tratamiento que se les ha asignado» (justificación del juicio).

Así pues, para la Cochrane, afirmar que una característica determinada puede generar sesgo requiere tres elementos fundamentales: identificar el mecanismo exclusivo por el que se introduciría, descartar otros posibles mecanismos y probar empíricamente de que ese mecanismo provoca un error sistemático. Examinamos a continuación el origen del concepto de «sesgo de financiación», la evidencia que lo sustenta y los mecanismos propuestos, para analizar después críticamente su existencia desde la perspectiva de la metodología Cochrane.

4. ¿Hay realmente un sesgo de financiación?

Teniendo en cuenta la definición de sesgo en investigación clínica, volvamos ahora sobre el concepto de *sesgo de financiación*. Pese a la contundencia de sus partidarios, el concepto no cuenta con una definición de consenso y ninguna de las propuestas que examinábamos al final de la sección segunda cuenta con apoyo mayoritario en la comunidad biomédica. A pesar de que la influencia de la financiación de la industria en la investigación clínica está ampliamente reconocida, lo que está en debate es que su efecto deba considerarse como un *sesgo* particular (Sterne, 2013; Binks, 2014; de Melo-Martin, 2019).

Hay, como se ha expuesto, evidencia clara y robusta sobre la asociación entre financiación por parte de la industria y mayor frecuencia de resultados favorables en la investigación clínica. Pero tanto la identificación del mecanismo a través del cual se introduce el error como la comprobación empírica que es ese mecanismo y no cualquier otro el que da lugar a resultados sistemáticamente diferentes está en debate. En primer lugar, los mecanismos que puntualmente pueden generar un sesgo a favor de la industria como los ya mencionados —selección de comparadores, etc.— son, de por sí, decisiones metodológicamente legítimas: el promotor de un estudio, siempre que cumpla las normas éticas de la investigación con humanos y la regulación que aplique en cada caso, puede elegir los comparadores y las variables que a su juicio capten mejor los efectos que desea medir, y hacerlo en uno u otro periodo de tiempo. De por sí, ninguna de estas decisiones provoca una desviación sistemática de la verdad al evaluar el efecto de un tratamiento. Puede que la información que genere el ECA no sea la más relevante para el médico que prescribirá el tratamiento o el seguro que lo financiará. Pero eso no quiere decir que el experimento esté sesgado (que ofrezca un resultado erróneo), simplemente obedece a otros intereses. Dada la calidad metodológica de los ECA patrocinados por la industria, si ese mismo experimento, con idénticas características, se realizase con fondos públicos, sus resultados serían probablemente idénticos. Más adelante se presentará evidencia de ello.

Por otro lado, los partidarios de la existencia del sesgo de financiación suelen atribuirle a la industria cierta omnipotencia para imponer sus intereses en el diseño y ejecución de ECA regulatorios. El promotor de un ensayo seleccionaría libremente y sin ningún tipo de limitación las características del estudio que va a llevar a cabo para conseguir unos resultados que sobreestimen el efecto del fármaco. Sin embargo, no es así como se realizan los ensayos clínicos. Cualquier compañía que desee poner uno en marcha debe solicitar autorización previa a la Agencia reguladora del territorio donde lo realiza (la FDA en EEUU, la EMA en Europa). Es la Agencia quien, a partir del examen de la información requerida, de acuerdo con la normativa y a través de un proceso formal predefinido, decide que la realización del ensayo clínico «es aceptable», «es aceptable [...] pero sometida al cumplimiento de condiciones específicas que se detallarán en la conclusión» o bien «no es aceptable»[6]. La normativa establece no solo el proceso a seguir para obtener la autorización de realizar el ensayo, sino también cómo debe llevarse a cabo, cómo debe monitorizarse, qué tipo de información debe facilitarse, cómo se realiza el seguimiento y comunicación de efectos adversos, etc... Además, la información clínicamente relevante debe ser revisada y aprobada por un Comité Ético, un órgano independiente a la Agencia, de composición multidisciplinar, cuya finalidad es velar por la protección de los derechos, seguridad y bienestar de las personas que participan en la investigación biomédica, ofreciendo garantía pública de ello. Por lo tanto, la industria no decide en ningún caso unilateralmente las características de los ensayos que desea llevar a cabo, sino que debe acordarlas con el regulador —y, por más que se hable de su captura, la evidencia al respecto es controvertida (Carpenter & Moss, 2013).

Partidarios de una concepción holista del sesgo, como Sismondo, podrían replicar que la verdadera manipulación se produce a espaldas del regulador, una vez que la industria despliega su batería de compañías auxiliares para realizar, analizar y difundir la investigación, donde puede imponer sus intereses sin cortapisas. Sin embargo, todas estas compañías auxiliares no están ocultas al regulador, operando en tierra de nadie. Las CROs son entidades perfectamente reguladas[7] y cada vez es más frecuente su participación en proyectos de investigación financiados públicamente. En realidad, la presencia de las CROs está más relacionada con la magnitud del proyecto que con la fuente de financiación. Sería posible citar numerosos casos en los que investigación financiada públicamente utiliza los servicios de una CRO, pero sirvan dos como ejemplo: en Estados Unidos, el estudio del *National Institute of Allergy and Infectious Diseases* (NIAID) para evaluar la eficacia de tocilizumab para reducir el número de rechazos en pacientes trasplantados de corazón está financiado exclusivamente con fondos públicos y, sin embargo cuenta con la participación de PPD, una de las CROs internacionales más importantes[8]; en Europa, el estudio I3LUNG, financiado exclusivamente por la Unión Europea y que tiene como objetivo promover el tratamiento individualizado en el cáncer de pulmón no microcítico (CPNMa), cuenta con la participación de MEDSIS, también una CRO (Prelaj *et al.*, 2023). La complejidad y grado de especialización que supone poner en marcha, monitorizar y coordinar un ensayo clínico que se lleva a cabo con pacientes localizados en múltiples centros, que generará una enorme cantidad de datos que requerirá de complejos análisis estadísticos, para el que será necesario elaborar diferentes publicaciones y en el que se deberán coordinar numerosos investigadores situados en lugares diferentes —todas ellas características cada vez más habituales— hacen que las CROs (con los servicios de operaciones clínicas, bioestadística, redacción médica, farmacovigilancia, gestión de publicaciones, etc.) se hayan convertido en un elemento fundamental de apoyo a la investigación clínica del siglo XXI, tanto pública como privada. Las CROs y las actividades que realizan no son un elemento oculto a la comunidad médica a través del cual la industria farmacéutica manipula y sesga la investigación, sino un actor más, reconocido y reconocible, que participa del proceso de acuerdo a la regulación existente, sea de financiación pública o privada.

---

6 En el caso de la Unión Europea, la realización de ensayos clínicos con medicamentos de uso humano se rige por el Reglamento (UE) No 536/2014. El reglamento requiere que en la solicitud de aprobación el promotor presente, entre otros: protocolo del estudio, manual del investigador, documentación sobre el cumplimiento de las normas de correcta fabricación del medicamento, información completa sobre las característica del medicamento en investigación (químicas y farmacológicas), información a los sujetos de ensayo, formulario de consentimiento informado y procedimiento de consentimiento informado, idoneidad del investigador, idoneidad de las instalaciones y prueba de la cobertura de seguro o de indemnización.

7 El Título 21 del Código de Regulaciones Federales de los Estados Unidos («Code of Federal Regulations CFR 21, 2022»), en el apartado 312.3, «Definiciones e interpretaciones», define qué considera la FDA una CRO: «Contract research organization means a person that assumes, as an independent contractor with the sponsor, one or more of the obligations of a sponsor, e.g. design of a protocol, selection or monitoring of investigators, evaluation of reports, and preparation of materials to be submitted to the Food and Drug Administration». Véase: https://www.ecfr.gov/current/title-21/chapter-I/subchapter-D/part-312 (consultado el 13/11/2024).

8 Véase «Tocilizumab in Heart Transplant» en https://clinicaltrials.gov/study/NCT03644667 (consultado el 13/11/2024).

Por otra parte, si bien el metanálisis de Lundh y colaboradores (2017) halla asociación entre financiación por la industria y mayor frecuencia de resultados favorables al promotor, eso no demuestra empíricamente que el patrocinio privado de la investigación introduzca un error sistemático. En investigación clínica, probar de modo concluyente la existencia del sesgo de financiación requeriría aportar evidencia de que, en las mismas condiciones de estudio, un tratamiento determinado tiene una eficacia diferente según la fuente de financiación. Es decir, que un fármaco X es más eficaz cuando la medición de su efecto se financia desde la industria que cuando se financia públicamente. Esto demostraría la existencia de un error sistemático provocado por el patrocinio de la industria farmacéutica y no por cualquier otra causa. Para ello sería preciso realizar una comparación directa, algo así como un ensayo clínico aleatorizado en el que la intervención que se evaluara fuera la «financiación», donde un grupo fuera financiado por la industria y otro públicamente. Ante las dificultades metodológicas evidentes de llevar a cabo un estudio de este tipo, una posibilidad alternativa es recurrir a evidencia indirecta: replicar con fondos públicos ensayos anteriormente realizados con patrocinio privado para comprobar si efectivamente se produce una desviación sistemática entre sus resultados. Si existe el sesgo de financiación es razonable esperar que, *ceteris paribus*, un ECA dará mejores resultados para un fármaco cuando lo financie la compañía que lo comercializa que cualquier patrocinador público. A pesar de que este tipo de comparación no es una prioridad para las agencias públicas, hay algún ejemplo que permite examinar qué ocurre en estos casos.

Ranibizumab (de nombre comercial Lucentis, propiedad de Novartis) es un anticuerpo monoclonal bloqueador del Factor de Crecimiento Endotelial Vascular (VEGF, por su acrónimo en inglés) que fue aprobado en 2006 por la FDA y en 2007 por la EMA para el tratamiento de la Degeneración Macular Asociada a la Edad (DMAE). Su similitud molecular y en el mecanismo de acción con un fármaco ya existente, bevacizumab (de nombre comercial Avastin, propiedad de Roche), que se había aprobado dos años antes por la FDA para el tratamiento de diferentes tipos de cáncer, generó un intenso debate. En la práctica clínica, antes de la comercialización de ranibizumab algunos oftalmólogos habían empezado a utilizar bevacizumab para el tratamiento de la DMAE. Aunque lo hacían fuera de indicación (no estaba autorizado para ese uso), la aprobación de ranibizumab para la DMAE consolidó y extendió esta práctica. El motivo era simple: siendo moléculas similares (ranibizumab era la fracción Fab aislada de bevacizumab) y teniendo el mismo mecanismo de acción (ambos eran antiVEGF), el precio de bevacizumab era considerablemente inferior al de ranibizumab[9]. Y si bien una parte de la comunidad médica mantenía que las diferencias moléculo-estructurales entre ambas las convertían en fármacos completamente diferentes[10], su uso en la práctica clínica obtenía resultados de eficacia similares. Finalmente, en el año 2008 un grupo de investigadores independiente de la Universidad de Pensilvania, agrupados bajo el nombre de *Comparison of Age-Related Macular Degeneration Treatments Trials (CATT) Research Group*, puso en marcha un ensayo clínico multicéntrico con más de 1.100 pacientes, financiado enteramente por el *National Eye Institute* de los Estados Unidos, con el objetivo de comparar la eficacia y seguridad de bevacizumab y ranibizumab en el tratamiento de la DMAE y aclarar así finalmente si eran fármacos que obtenían resultados similares.

Novartis obtuvo la autorización de ranibizumab con los resultados de dos ensayos clínicos controlados con placebo (MARINA (Rosenfeld *et al.*, 2006) y ANCHOR (Brown *et al.*, 2006)) en los que demostraba la eficacia del fármaco para la DMAE. En el estudio independiente, los investigadores del *CATT Research Group* realizaron también un análisis específico sobre los pacientes que habían recibido ranibizumab, aplicando los criterios de inclusión de MARINA y ANCHOR. Con ello buscaban conseguir la máxima similitud entre sus pacientes y los pacientes de los ensayos clínicos financiados por la compañía farmacéutica. En su presentación de resultados (CATT Research Group, 2011) afirmaron que los aumentos de agudeza visual obtenidos en el estudio CATT y en MARINA y ANCHOR fueron similares[11]. Es

---

9 La relación de precio entre ranibizumab y bevacizumab era en 2007 de 39:1, basándose en el precio en EE.UU: 1.950 dólares por inyección para ranibizumab y 50 dólares para bevacizumab (Raftery *et al.*, 2007).

10 Entre otras, que al ser ranibizumab más pequeño resultaba 140 veces más específico en su unión al VEGF.

11 «Los aumentos medios de agudeza visual de los dos subgrupos fueron similares a los de los ensayos pivotales. Estos aumentos fueron de 9,8 letras en el CATT frente a 7,2 letras en el ensayo (...) del anticuerpo anti-VEGF ranibizumab en el tratamiento de la DMAE Neovascular (MARINA; Número de ClinicalTrials.gov, NCT00056836) y 10,8 letras en el CATT frente a 11,3 letras en el anticuerpo anti-VEGF para

decir, *cuando investigadores independientes realizaron un análisis con el objetivo específico de hallar diferencias entre los resultados de eficacia que habían obtenido los estudios financiados por la industria farmacéutica y los resultados del estudio financiado públicamente no encontraron ninguna significativa*. Al estar los estudios MARINA y ANCHOR patrocinados por Novartis, debería haberse producido el sesgo de financiación, es decir, debería haberse producido un error sistemático en los resultados sobreestimando el efecto del fármaco. Pero el estudio independiente mostró que no se había producido tal sobreestimación. Es decir, que la financiación de la farmacéutica no había modificado los resultados a su favor: no había sesgo de financiación.

5. *Conclusión*

La existencia de un *sesgo de financiación* que genera un error sistemático consecuencia de la financiación por la industria farmacéutica requiere, según criterio de la Cochrane, la identificación y descripción del mecanismo por el que actúa y la comprobación empírica de su efecto. Como se ha expuesto, hasta la fecha ni se ha identificado de forma inequívoca el mecanismo exclusivo por el cual actúa la financiación privada ni se ha aportado evidencia robusta de que esta financiación, *per se,* introduzca un error sistemático en los resultados. Además, existe evidencia indirecta de que cuando los efectos de un fármaco se evalúan, en condiciones equivalentes, a través de estudios financiados por la industria y estudios financiados públicamente, los resultados de eficacia y seguridad no presentan diferencias significativas, lo que cuestiona que la influencia de la industria pueda ser considerada como generadora de error sistemático.

Ante la afirmación de que no existe un *sesgo de financiación* pueden plantearse dos posibles objeciones. En primer lugar, se podría cuestionar la definición del concepto de *sesgo* que utiliza la Cochrane, abogando por alternativas más amplias y menos específicas. No obstante, a pesar de que el término *sesgo* es polisémico y a veces se utiliza de forma diferente en ciencia y en filosofía (Wilholt, 2009, p. 92), es innegable que el concepto implica de forma general la introducción de un error que desvía el proceso científico de los resultados verdaderos. Sin duda, los filósofos pueden jugar un importante papel en la evaluación y mejora de la ciencia patrocinada por la industria (Holman & Elliot, 2018), pero si desde la filosofía y la sociología de la ciencia se desea reflexionar sobre la práctica científica de forma que interpele realmente a sus actores, resulta necesario utilizar los términos científicos con el significado y uso que le dan los propios científicos. Como mantiene Hull (1998, p. 298) «[l]o que los científicos tienen que decir sobre esos términos tiene cierta prioridad sobre la opinión de cualquier otro, incluyendo los filósofos [...] el árbitro último en estos casos es el uso». Y no hay duda que en el análisis de sesgos en la investigación clínica, la Cochrane es el referente para investigadores y médicos. En segundo lugar, podría plantearse que de no existir el sesgo de financiación quedaría sin explicar cómo se produce la asociación empíricamente demostrada entre mayor frecuencia de resultados favorables y estudios financiados por la industria. Sobre este punto se han propuesto diferentes posibles explicaciones alternativas (ver, por ejemplo, Lundh *et al.*, 2017), y lo más probable sea que no haya una única causa común a todos los estudios. En esta línea se encuentra la propuesta de De Melo-Martín (2019). Para la autora, cuando la industria introduce errores sistemáticos en la investigación clínica lo hace a través de mecanismos concretos particulares en cada caso. Es decir, la financiación no sería *per se* lo que provocaría el sesgo, sino el motor que podría activar algunos de los diferentes mecanismos posibles a través de los cuales se alteran resultados. Será preciso identificar en cada caso el mecanismo y corregirlo específicamente. También Krimsky (2013) señala que la financiación sería tan solo un signo de la posible presencia de factores que pueden influir en los resultados de investigación, factores que deben ser investigados en cada caso y para los que no debe recurrirse automáticamente a la existencia de *sesgo* como explicación. Contra la visión holista defendida por Sismondo, no tendríamos, por tanto, un sesgo de financiación, sino múltiples sesgos particulares.

el tratamiento de la neovascularización coroidea predominantemente clásica en la DMAE (ANCHOR, NCT00061594)» (CATT Research Group, 2011, p. 1902-3).

La clave del éxito de la metodología Cochrane ha consistido en trabajar caso a caso sobre sesgos particulares y sus dispositivos de corrección, identificando ambos con el consenso de la comunidad biomédica. La investigación biomédica se desarrolla entre conflictos de intereses: las compañías farmacéuticas compiten entre sí por llevar sus tratamientos a los mercados; los investigadores compiten por el reconocimiento social y las oportunidades que les ofrece su descubrimiento; los pacientes de los ECA rara vez son indiferentes a los medicamentos que se prueban y sus propias preferencias pueden tener influencia en el resultado. La *RoB* sirve como indicador para saber si en cada ECA se han controlado todos los posibles mecanismos conocidos a través de los cuales esos conflictos de intereses pueden afectar al resultado y facilitar el acuerdo de los competidores sobre el resultado (Teira, 2016). Declarar el patrocinio industrial como fuente de sesgo no facilita el acuerdo, sino que lo complica, como muestra la radicalidad de las medidas propuestas por Sismondo y otros críticos: si toda investigación financiada por la industria es sospechosa, no se debería creer, por sistema, la mayor parte de los estudios publicados en las revistas médicas más prestigiosas –conclusión que críticos como Gotzsche no tienen reparo en aceptar.

Es posible que la medicina clínica necesite ECA diferentes a los patrocinados por la industria, pero eso no convierte en ilegítimos a estos últimos. Igual que con el resto de sesgos conocidos, la existencia de un sesgo de financiación en la investigación clínica requiere, además de identificar y describir el mecanismo concreto y exclusivo que lo produce, evidencia empírica de que es ese mecanismo y no otro el que introduce un error sistemático en los resultados convirtiéndolos en no fiables. Y dado que las propuestas existentes hasta la fecha no han probado que el efecto de un fármaco sea diferente según en los estudios se den o no los mecanismos postulados, resulta más razonable identificar y corregir los posibles errores de acuerdo a los estándares de la investigación clínica: caso a caso y evitando su estigmatización según el origen, como si de un pecado original se tratara.

Agradecimientos

El origen del artículo se encuentra en las numerosas conversaciones mantenidas con David Teira, de quien debo destacar su absoluta disponibilidad, ponderado criterio y acertada crítica. La paciencia y buena disposición de las especialistas en investigación clínica Montse Balañà, Griselda Martrat e Irma Roig ha sido fundamental para su elaboración: sin sus respuestas a mis numerosas consultas hubiera sido todo mucho más difícil. Manuela Fernández Pinto tuvo la amabilidad de leer una versión preliminar y aportar comentarios relevantes. Finalmente, dos revisores anónimos propusieron mejoras sustanciales a su contenido. A todos ellos mi agradecimiento sincero.

Conflictos de interés

Francesc Roig-Loscertales ha trabajado en la industria farmacéutica y en la universidad.

REFERENCIAS

Als-Nielsen, B., Chen, W., Gluud, C., Kjaergard, LL. (2003). Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 290(7), 921-928.

Angell, M. (2005). *The truth about the drug companies: How they deceive us and what to do about it*. Random House.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D., Stroup, D.F. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 276(8), 637-639.

Bekelman, J. E., Li, Y., Gross, C. P. (2003). Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 289 (4), 454-465.

Bero, LA. (2013). Why the Cochrane Risk of Bias Tool Should Include Funding Source as a Standard Item. *Cochrane Database of Systematic Reviews* Issue 12. Art. No.: ED000075.

Bero, L. (2018). More journals should have conflict of interest policies as strict as Cochrane. *BMJ Opinion* https://blogs.bmj.com/bmj/2018/11/12/lisa-bero-more-journals-should-have-conflict-of-interest-policies-as-strict-as-cochrane/ (consultado el 13/11/2024).

Binks, M. (2014). Judge the science, not the funding source. *International Journal of Obesity* 38(5), 625.

Bone, H.G., Kiel, D.P., Lindsay, R.S., Lewiecki, E.M., Bolognese, M.A., Leary, E.T., *et al.* (2007). Effects of atorvastatin on bone in postmenopausal women with dyslipidemia: A double-blind, placebo-controlled, dose-ranging trial. *Journal of Clinical Endocrinology and Metabolism* 92(12), 4671-4677.

Boutron, I., Page, M.J., Higgins, J.P.T., Altman, D.G., Lundh, A., Hróbjartsson, A. (2024). Considering bias and conflicts of interest among the included studies [last updated August 2022]. In Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A. (eds.). *Cochrane Handbook for Systematic Reviews of Interventions version 6.5. Cochrane*.

Brown, D.M., Kaiser, P.K., Michels, M., *et al.* (2006). Ranibizumab versus verteporfin for neovascular age-related macular degeneration. *New England Journal of Medicine* 355, 1432-1444.

Campbell, D.T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* 54, 297-312.

Carpenter, D., Moss, D.A. (2013). *Preventing regulatory capture. Special interest influence and how to limit it*. Cambridge University Press.

Chalmers, T. C., Matta, R. J., Smith, H., & Kunzler, A.-M. (1977). Evidence favoring the use of anticoagulants in the hospital phase of acute myocardial infarction. *New England Journal of Medicine* 297(20), 1091-1096.

Claridge, J. A., Fabian, T. C. (2005). History and development of evidence-based medicine. *World Journal of Surgery* 29(5), 547–553.

Cumpston, M., Flemyng, E., Thomas, J., Higgins, J.P.T., Deeks, J.J., Clarke, M.J. (2024). Introduction. In Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A. (eds.). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5.

de Melo-Martín, I. (2019). The commercialization of the biomedical sciences: (mis)understanding bias. *History and Philosophy of the Life Sciences* 41(3), 34.

De Vito, N. J., Bacon, S., Goldacre, B. (2020). Compliance with legal requirement to report clinical trial results on ClinicalTrials.gov: a cohort study. *Lancet* 395, 361-369.

Dechartres, A., Boutron, I., Trinquart, L., Charles, P., Ravaud, P. (2011). Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. *Annals of Internal Medicine* 155, 39-51.

Dechartres, A., Trinquart, L., Boutron, I., Ravaud, P. (2013). Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ* 346, f2304.

Dechartres, A., Trinquart, L., Faber, T., Ravaud, P. (2016). Empirical evaluation of which trial characteristics are associated with treatment effect estimates. *Journal of Clinical Epidemiology* 77, 24-37.

Djulbegovic, B., Lacevic, M., Cantor, A., Fields, K.K., Bennett, C.L., Adams, J.R., *et al.* (2000). The uncertainty principle and industrys ponsored research. *Lancet* 356(9230), 635-638.

Djulbegovic, B., Kumar, A., Miladinovic, B., Reljic, T., Galeb, S., Mhaskar, A., *et al.* (2013). Treatment success in cancer: industry compared to publicly sponsored randomized controlled trials. *PLoS One* 8(3), e58711.

Dunn, A.G., Mandl, K.D., Coiera, E., Bourgeois, F.T. (2013). The effects of industry sponsorship on comparator selection in trial registrations for neuropsychiatric conditions in children. *PLoS One* 8(12), e84951.

Estellat, C., Ravaud, P. (2012). Lack of head-to-head trials and fair control arms: randomized controlled trials of biologic treatment for rheumatoid arthritis. *Archives of Internal Medicine* 172(3), 237-44.

Fernández Pinto, M. (2023) Methodological and cognitive biases in science: Issues for current research and ways to counteract them. *Perspectives on Science* 31(5), 535-554.

Flacco, M.E., Manzoli, L., Boccia, S., Capasso, L., Aleksovska, K., Rosso, A., Scaioli, G., *et al.* (2015). Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor. *Journal of Clinical Epidemiology* 68(7), 811-820.

Flemyng, E., Moore, T.H., Boutron, I., *et al.* (2023). Using Risk of Bias 2 to assess results from randomised controlled trials: guidance from Cochrane. *BMJ Evidence-Based Medicine* 28, 260-266.

Gazendam, A.M., Slawaska-Eng, D., Nucci, N., Bhatt, O., Ghert, M. (2022). The impact of industry funding on randomized controlled trials of biologic therapies. *Medicines* 9(18).

Goldacre, B., DeVito, N. J., Heneghan, C., Irving, F., Bacon, S., Fleminger, J., Curtis, H. (2018). Compliance with requirement to report results on the EU Clinical Trials Register: cohort study and web resource. *BMJ (Clinical research edition)* 362, k3218.

Gøtzsche, P.C. (2013). *Deadly medicines and organised crime: how big pharma has corrupted healthcare*. Radcliffe Publishing.

Hemminki, H. (1980). Study of information submitted by drug companies to licensing authorities. *BMJ* 280, 833-836.

Higgins, J.P., Altman, D.G., Gotzsche, P.C., Juni, P., Moher, D., Oxman, A.D., *et al.* (2011). The Cochrane Collaboration's Tool for assessing risk of bias in randomised trials. *BMJ* 343, d5928.

Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A. (eds.) (2024). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5. Cochrane.

Holman, B., Elliott, K.C. (2018). The promise and perils of industry-funded science. *Philosophy Compass* 13, e12544.

Hull, D.L. (1998). *Science as a process. An evolutionary account of the social and conceptual development of science*. The University of Chicago Press.

Hulley, S., Grady. D., Bush, T. *et al.* (1998). Randomized trial of estrogen plus progestin for secondary prevention of coronary artery disease in postmenopausal women. *JAMA* 280, 605-613.

Jefferson, T. (2020). Sponsorship bias in clinical trials: growing menace or dawning realisation? *Journal of the Royal Society of Medicine* 113(4), 148-157.

Johansen, H.K., Gøtzsche, P.C. (1999). Problems in the design and reporting of trials of antifungal agents encountered during meta-analysis. *JAMA* 282(18), 1752-1759.

Jüni, P., Altman, D.G., Egger, M. (2001). Systematic reviews in health care: assessing the quality of controlled clinical trials. *BMJ* 23, 42-46.

Jüni P., Dieppe, P.A. (2002). Are selective COX 2 inhibitors superior to traditional nonsteroidal anti-inflammatory drugs? *BMJ* 324, 1287.

Jüni, P., Witschi, A., Bloch, R., *et al.* (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 282, 1054-1060.

Katz, K.A., Karlawish, J.H., Chiang, D.S., Bognet, R.A., Propert, K.J., Margolis, D.J. (2006). Prevalence and factors associated with use of placebo control groups in randomized controlled trials in psoriasis: a cross-sectional study. *Journal of the American Academy of Dermatology* 55(5), 814-822.

Krimsky, S. (2003). *Science in the private interest: Has the lure of profits corrupted biomedical research?* Rowman & Littlefield Publishers.

Krimsky, S. (2013). Do financial conflicts of inteterest bias research? An inquiry into the "funding effect" hypothesis. *Science, Technology, & Human Values*, 8(4), 566-587.

Laine, C., De Angelis, C., Delamothe, T., Drazen, J.M., Frizelle, F.A., Haug, C., Hébert, P.C., Horton, R., Kotzin, S., Marusic, A., Sahni, P., Schroeder, T.V., Sox, H.C., Van der Weyden, M.B., Verheugt, F.W. (2007). Clinical trial registration: looking back and moving ahead. *Annals of Internal Medicine* 147(4), 275-277.

Lathyris, D.N., Patsopoulos, N.A., Salanti, G., Ioannidis, J.P. (2010). Industry sponsorship and selection of comparators in randomized clinical trials. *European Journal of Clinical Investigation* 40(2), 172-82.

Leefmann, J. (2021). How to assess the epistemic wrongness of sponsorship bias? The case of manufactured certainty. *Frontiers in research metrics and analitics* 6, 599909.

Lexchin, J., Bero, L. A., Djulbegovic, B., Clark, O. (2003). Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. *BMJ (Clinical research edition)* 326(7400), 1167–1170.

Lundh, A., Gøtzsche, P.C. (2008). Recommendations by Cochrane Review Groups for assessment of the risk of bias in studies. *BMC Medical Research Methodology* 8, 22.

Lundh, A., Sismondo, S., Lexchin, J., Busuioc, O.A., Bero L. (2012). Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews* 12, MR000033.

Lundh,A., Lexchin, J., Mintzes,B., Schroll, J.B., Bero, L. (2017). Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews* 2(2), MR000033.

Page, M.J., Higgins, J.P.T., Clayton, G., Sterne, J.A.C., Hróbjartsson, A., Savović, J. (2016). Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. *PLoS One* 11, 7.

Piller, C. (2020). Transparency on trial. *Science* 367(6475), 240-243.

Prelaj, A., Ganzinelli, M., Trovo', F., Roisman, L.C., Pedrocchi, A.L.G., Kosta, S., Restelli, M., Ambrosini, E., Broggini, M., Pravettoni, G., Monzani, D., Nuara, A., Amat, R., Spathas, N., Willis, M., Pearson, A., Dolezal, J., Mazzeo, L., Sangaletti, S., Correa, A.M., Aguaron, A., Watermann, I., Popa, C., Raimondi, G., Triulzi, T., Steurer, S., Lo Russo, G., Linardou, H., Peled, N., Felip, E., Reck, M., Garassino, M.C. (2023). The EU-funded I^{3}LUNG Project: Integrative Science, Intelligent Data Platform for Individualized LUNG Cancer Care With Immunotherapy. *Clinical Lung Cancer*. 24(4), 381-387.

Raftery, J., Clegg, A., Jones, J., Tan, S.C. (2007). Lotery A. Ranibizumab (Lucentis) versus bevacizumab (Avastin): modelling cost effectiveness. *British Journal of Ophthalmology* 91(9), 1244-1246.

Resnik, D. B., Elliott, K. C. (2013). Taking financial relationships into account when assessing research. *Accountability in Research* 20(3), 184-205.

Reutlinger, A. (2020). What is epistemically wrong with research affected by sponsorship bias? The evidential account. *European Journal for Philosophy of Science* 10(2), 1-26.

Rochon, P.A., Gurwitz, J.H., Simms, R.W., Fortin, P.R., Felson, D.T., Minaker, K.L., *et al.* (1994). A study of manufacturer-supported trials of nonsteroidal anti-inflammatory drugs in the treatment of arthritis. *Archives of Internal Medicine* 154(2), 157-63.

Rosenfeld, P.J., Brown, D.M., Heier, J.S., *et al.* (2006). Ranibizumab for neovascular age-related macular degeneration. *New England Journal of Medicine* 355, 1419-1431.

Safer, D.J. (2002). Design and reporting modifications in industry sponsored comparative psychopharmacology trials. *Journal of Nervous and Mental Disease* 190, 583-592.

Savović, J., Jones, H.E., Altman, D.G., Harris, R.J., Jüni, P., Pildal, J., Als-Nielsen, B., Balk, E.M., Gluud, C., Gluud, L.L., Ioannidis, J.P.A., Schulz, K.F., Beynon, R., Welton, N.J., Wood, L., Moher, D., Deeks, J.J., Sterne, J.A.C. (2012). Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Annals of Internal Medicine* 157, 429-438.

Schatzkin, A., Lanza, E., Corle, D. *et al.* for the Polyp Prevention Trial Study Group (2000). Lack of effect of a low-fat, high-fiber diet on the recurrence of colorectal adenomas. *New England Journal of Medicine* 342, 1149-1155.

Schott, G., Pachl, H., Limbach, U., Gundert-Remy, U., Ludwig, W.D., Lieb, K. (2010). The financing of drug trials by pharmaceutical companies and its consequences, part 1: a qualitative, systematic review of the literature on possible influences on the findings, protocols, and quality of drug trials. *Deutsches Ärzteblatt International* 107 (16), 279.

Schulz, K.F., Chalmers, I., Hayes, R.J., Altman, D.G. (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273, 408-412.

Sinyor, M., Schaffer, A., Smart, K.A., Levitt, A.J., Lanctot, K.L., Grysman, N.H. (2012) Sponsorship, antidepressant dose, and outcome in major depressive disorder: meta-analysis of randomized controlled trials. *Journal of Clinical Psychiatry* 73(2), e277-87.

Sismondo, S. (2008). How pharmaceutical industry funding affects trial outcomes: Causal structures and responses. *Social Science & Medicine* 66(9), 1909-1914.

Sismondo, S. (2021). Epistemic corruption, the pharmaceutical industry, and the body of medical science. *Frontiers in research metrics and analytics* 6, 614013.

Solomon, M. (2020). After disclosure. In LaCaze, A., Osimani, B. (eds.) *Uncertainty in Pharmacology* (439-450). Springer.

Sollmann, T. (1917). The crucial test of therapeutic evidence. *JAMA* 69, 198-210.

Sterne, J.A.C. (2013). Why the cochrane risk of bias tool should not include funding source as a standard item. *Cochrane Database of Systematic Reviews* 12, ED000076.

Sterne, J.A.C., Savović, J., Page, M.J., *et al.* (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 366, 4898.

Taylor, F., Huffman, M.D., Macedo, A.F., Moore, T.H.M., Burke, M., Davey Smith, G., Ward, K., Ebrahim, S., Gay, H.C. (2013). Statins for the primary prevention of cardiovascular disease. *Cochrane Database of Systematic Reviews* 1, CD004816.

Teira, D. (2016). Debiasing methods and the acceptability of experimental outcomes. *Perspectives on Science* 24(6), 722-743.

The Alpha Tocopherol Beta Carotene Cancer Prevention Study Group. (1994). The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *New England Journal of Medicine* 330, 1029-1035.

Tse, T., Zarin, D.A. (2009). Clinical trial registration and results reporting: ClinicalTrials.gov and FDAAA. *FDLI Update* 1, 18-22.

Vandenbroucke, J.P., von Elm, E., Altman, D.G., Gøtzsche, P.C., Mulrow, C.D., Pocock, S.J., Poole, C., Schlesselman, J.J., Egger, M. (2007). STROBE initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Annals of Internal Medicine* 147(8), W163-194.

Vedula, S.S., Bero, L., Scherer, R.W., Dickersin, K. (2009). Outcome reporting in industry-sponsored trials of Gabapentin for off-label use. *New England Journal of Medicine* 361(20), 1963-1971.

Vetter, T.R., Mascha, E.J. (2017). Bias, confounding, and interaction: lions and tigers, and bears, Oh my! *Anesthesia & Analgesia* 125(3), 1042-1048.

Viergever, R.F., Li, K. (2015). Trends in global clinical trial registration: an analysis of numbers of registered clinical trials in different parts of the world from 2004 to 2013. *BMJ Open* 5, e008932.

Wilholt, T. (2009). Bias and values in scientific research. *Studies in History and Philosophy of Science Part A* 40(1): 92-101.

Wood, L., Egger, M., Gluud, L.L., Schulz, K., Jüni, P., Altman, D.G., *et al.* (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 336, 601-605.

Yusuf, S., Dagenais, G., Pogue, J., Bosch, J., Sleight, P., Heart Outcomes Prevention Evaluation Study Investigators (2000). Vitamin E supplementation and cardiovascular events in high-risk patients for the Heart Outcomes Prevention Study Investigators. *New England Journal of Medicine* 342, 154-160.

Francesc Roig-Loscertales realiza su investigación en el Departamento de Lógica, Historia y Filosofía de la ciencia de la UNED y centra su ámbito de interés en el análisis de la actividad de la industria farmacéutica.

Dirección postal: Dpto. de Lógica, Historia y Filosofía de la ciencia, UNED. Paseo de Senda del Rey 7, 28040 Madrid, España. E-mail: froig9@alumno.uned.es – ORCID: 0000-0002-3585-7536

Theoria

An International Journal for Theory, History and Foundations of Science

'Ο Θεὸς ἀριθμετίζει
THEORIA
ISSN 0495-4548 – eISSN 2171-679X

Summary

Volume 41/1, January 2026, pp. 1-109

Articles

Luis Estrada-González, Christian Romero-Rodríguez (Institute for Philosophical Research, UNAM), «How we learned to stop worrying and love tonk», *Theoria*, 2026, 41/1, 5-21

> Belnap highlighted the role of Transitivity in Prior's triviality proof involving *tonk*, but a non-trivial, non-transitive logic with *tonk* was never developed until Cook's proposal with four interpretations and a disjunctive consequence relation. We improve on that proposal: we show that only three interpretations suffice and that a non-disjunctive consequence relation is not required.

Leonardo Flamini (University of Johannesburg), «On asking», *Theoria*, 2026, 41/1, 22-35

> Against Watson's information-seeking account of questions, I defend an aim-constitutivist view drawing on Searle's conception of the illocutionary point of directives. I also refine sincerity conditions via Friedman's notion of inquiring attitudes and show how questions can be recognised not only as speech acts but also as mental states.

Sergio Cermeño-Aínsa (University of Girona), «Perceptual reference, object files and Molyneux's Question», *Theoria*, 2026, 41/1, 36-56

> This paper defends a solution to the Molyneux Question grounded on what I call the *object files strategy*. Roughly, if tactile information is referentially encoded and amodally stored in object files, then the category-specific contents of R(vision) and R(touch) are in some way propositionally (conceptually) linked, and therefore a newly sighted might recognize tactilely familiar shapes by sight alone.

Peter J. Riggs (Australian National University), «Time dilation and rates of the passage of time», *Theoria*, 2026, 41/1, 57-70

> There has not been any substantial progress made on the issue of the rate of the passage of time. This article proposes a way to make such progress and draws a significant conclusion about the nature of time.

Eduard Dàvila (University of Valladolid), «A critical analysis of cosmological typicality and the anthropic principle», *Theoria*, 2026, 41/1, 71-86

This paper critically examines Stephen Weinberg's anthropic prediction of the cosmological constant. It analyzes three methodological components of his argument: the probability measure, the conditionalization linking observers to galaxy formation, and the assumption of typicality. I argue that the predictive force of the model depends on substantive commitments about prior distributions and the treatment of self-locating uncertainty.

Francesc Roig-Loscertales (UNED), «¿Hay un sesgo de financiación en la investigación biomédica?», *Theoria*, 2026, 41/1, 87-102

The article critically examines the notion of "funding bias" in the pharmaceutical industry sponsored clinical research. Based on the definition of bias established by the Cochrane methodology, it argues that the possible influence of commercial interests does not meet the criteria required to be properly considered a bias within the field of clinical research.

Referees for Theoria
Informantes de Theoria
(2023-2024 / 2024-2025)

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

Referees for THEORIA (2023-2024 / 2024-2025)
Informantes de THEORIA (2023-2024 / 2024-2025)

The editors gratefully acknowledge the contribution of the referees listed below. They all reviewed one or more manuscripts submitted between 2023-2024 and 2024-2025. Their reports helped the members of our Editorial Board in making their final decisions.

El equipo editor agradece la colaboración de las personas evaluadoras que se detallan debajo. Cada una de ellas emitió informes sobre uno o varios manuscritos recibidos en los periodos 2023-2024 y 2024-2025 . Sus informes sirvieron para fundamentar las decisiones del Consejo editorial sobre su publicación.

Martín Abreu Zavaleta
Joshua J. Alexander
Valia Allori
Manuel Almagro
Ramón Alvarado
Miguel Álvarez Lisboa
Luvell Anderson
Gloria Andrada de Gregorio
Mattia Andreoletti
Rani Lill Anjum
Jamin Asay
Brandon James Ashby
Paul Audi
Gunnar Babcock
Matteo Baggio
Sergio Balari
David Barack
Lorenzo Baravalle
Áxel Barceló
Eduardo Barrio
Lucien Baumgartner
Claus Beisbart
Francesca Bellazzi
Karl Bergman
Lilian Bermejo
Reinaldo Bernal
Joan Bertrán San Millán
Alexandre Billon
Brandon Boesch
Florian Boge
Paolo Bonardi
Marcelino Botín
David Bourget
Mark Bowker
Vojislav Božičković
Adam Bradley
Liam Bright
Jelle Bruineberg
Wesley Buckwalter
David Builes
María Caamaño
Fabrizio Calzavarini
Fabrizio Calzavarini
Elisabeth Camp
Carlos Enrique Caorsi
Laura Caponetto
Massimiliano Carrara
Paula Casal
Alejandro Cassini
Dan Cavedon-Taylor
María Cerezo
Lee-Ann Chae
Amit Chaturvedi
Eddy Kemming Cheng
Enrico Cinti
Hasok Chang
Tony Cheng
Pablo Cobreros
Dimitri Coelho Mollo
Sam Coleman
Tim Crane
Lindsay Crawford
Adrian Currie
André Curtis-Trudel
Ciro De Florio
Gui Sanches de Oliveira
María de Ponte
Richard de Rozario
Cristina Corredor
Ana Cuevas Badallo
André Curtis-Trudel
Silvia De Toffoli
Remy Debes
Fabio Del Prete
Ophelia Deroy
Hugh Desmond
Nicola Di Stefano
Bogdan Dicher
Dennis Dieks
James DiFrisco
José Antonio Díez Calzada
Chao Ding
Ned Dobos
Krzysztof Dolega

John Dorsch
Louis Doulas
Zoe Drayson
Anna Drożdżowicz
Laura Duhau
Santiago Echeverri
Miguel Egler
Carlos Elías
Matthias Egg
Gretchen Ellefson
Luis Estrada
Marco Facchin
Marco Facchin
Boaz Faraday Schuman
Catia Faria
Matt Farr
Giulia Felappi
Matt Ferkany
Pablo Fernández Velasco
Gabriele Ferretti
Aldo Filomeno
Eugen Fischer
Sarah Fisher
Samuel C. Fletcher
Carolina Flores
Janet Folina
Keith Frankish
María José Frapolli
Doreen Fraser
James Fraser
Nir Fresco
Héctor Freytes
James Fritz
Karen Frost-Arnold
Naoya Fujikawa
Antonio Gaitán
Enrico Galli
Robert K García
Pío García
Manuel García Carpintero
María José García-Encinas
Manuel J. García-Pérez
José Ángel Gascón
Grzegorz Gaszczyk
Santiago Ginnobili
David Glick
Bernardo Gonçalves
Keran Gordon-Solmon
Vincent Grandjean
Aidan Gray
Alex Gregory
Daniel Gregory
Ronald Gruber
Bojana Grujičić
Alex Grzankowski
Francesco Guala
Mario Guenther
Abel Gustafson
Charles H. Pence
Nina Haket
Marta Halina
Peter Hanks
Margherita Harris
Joyce Havstad
Eliran Haziza
Paul Henne
Cassie Herbert
Manuel Heras Escribano
Tyler Hildebrand
Inês Hipólito
Adam Hochman
Federico Holik
David Hommen
Oscar Horta
Robert Howell
Fabian Hundertmark
Josh Hunt
Manuel J. García-Pérez
Carme Isern-Mas
Rasmus Jaksland
Christian Jendreiko
Ethan Jerzak
Juan Pablo Jorge
Zuzanna Jusińska
Nino Kadic
François Kammerer
Yannic Kappes
Megan Katsumi
Markku Keinänen
Eddy Kemming Cheng
Alex Kerr
David Kinney
Kepa Korta
Bas Kortenbach
Decio Krause
Uriah Kriegel
Elio La Rosa
Peter Langland-Hassan
Victoria Lavorerio
Dustin Lazarovici
Baptiste Le Bihan
Olivier Lemeire
Juliana Lima
Michelle Liu
Vasileios Livanios
Olimpia Lombardi
Guy Longworth
Cristian López
Lorenzo Lorenzetti
Natalia Luna
Victor J. Luque
Fabrizio Macagno
Lilith Mace
Josep Macià
Alice MacLachlan
Fiona Macpherson
Paolo Maffezioli
Pablo Magaña
P.D. Magnus
Filippo Mancini
Stephen Mann
Ben Martin
María Laura Martínez
María Martínez Ordaz
Fernando Martínez-Manrique
Carlo Martini
Claudio Masolo
Lucy Mason
Dana Matthiessen
Anna-Sofia Maurin
Tom McClelland
Kael McCormack
Kris McDaniel
Angela Mendelovici
José Jorge Mendoza
Tushar Menon
Joseph Metz
Michael Michael
Sofia Miguens
Marcin Miłkowski
Rapahël Millière
Garrett Mindt
Susana Monsó
Martin Montminy
Dario Mortini
Sebastiano Moruzzi
Cristian Moyano
Léna Mudry
Daniel Muñoz
Sebastian Murgueitio Ramirez
Adriana Murguía Lores
Alice Murphy
Hichem Naar

Jesús Navarro Reyes
Ilkka Niiniluoto
Ethan Nowak
Nancy Nuñez
Elias Okon
Andrea Onofri
Krisztina Orbán
Ismael Ordóñez Miguéns
Ezequiel Paez
Maria Pia Paganelli
Jacopo Pallagrosi
Michele Palmira
Graham Parsons
Matias Pasqualini
Ross Patrizio
Andrew Peet
Diana Pérez
José Antonio Pérez Escobar
Saúl Pérez González
Eduardo Pérez-Navarro
Francesca Pero
Bryan Pickel
Vicente Pico
David Pineda
Mihaela Popa
Jesús Portillo-Fernández
Nina Poth
Simon Prosser
Sophie R. Allen
Carlo Raineri
Chris Ranalli
Indrek Reiland
Andrew Reisner
Ivan Restovic
Collin Rice
Nicholas Rimell
Regina Rini
J. Brendan Ritchie
Alexander Roberts
Sebastian Rödl
Luke Roelofs
Ariel Roffé
Aida Roige
Luis Rosa
Lucas Rosenblatt
Anna-Mari Rusanen
Stéphanie Ruphy
Quentin Ruyant
Pablo Rychter
Thomas Ryckman
Juha Saatsi
Alexander Sager
Catharine Saint-Croix
Léa Salje
Gui Sanches de Oliveira
Glenda Satne
Susanna Schellenberg
Dietram. A. Scheufele
Raphael Scholl
Peter Schulte
Armin Schulz
Silvia Carolina Scotto
Miguel Ángel Sebastián
Johanna Seibt
Markus Seidel
Giuseppe Sergioli
Dunja Seselja
Dunja Šešelja
Hamid Seyedsayamdost
Oron Shagrir
Yafeng Shan
Ashley Shaw
Joshua Shepherd
Kenneth Silver
Ylwa Sjölin Wirling
Léna Soler
Aris Spanos
Katie Steele
Florian Steinberger
Justin Stysma
Javier Suárez
Brianne Suldovsky
Mieszko Tałasiewicz
Pilar Terrés
Justin Tiehen
Lisa Titus
Stephan Torre
Robert Trueman
Jonathan Tsou
Matthew Tugby
Joaquín Miguel Valdivielso Navarro
Íñigo Valero
Víctor M. Verdejo
Alfredo Vernazzani
Agustín Vicente
Laure Vieu
Massimiliano Vignolo
Cristina Villegas
Alberto Voltolini
Bryan W Roberts
Aja Watkins
Daniel Waxman
Sam Wilkinson
Malte Willer
Belle Wollesen
Benjamin Young
Nick Young
José Zalabardo
Carlos Zednik
Henrik Zinkernagel

https://doi.org/10.1387/theoria.27281

THEORIA

An International Journal for Theory, History and Foundations of Science

Vol. 41/1 • January 2026

Dep. Legal: SS - 820/85 ISSN 0495-4548 e-ISSN 2171-679X

Articles